Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*

CONTENTS

Vexierversuch: The Log Relationship Between Word-Frequency and Recognition Obtained in the Absence of Stimulus Words: I. GOLDIAMOND AND W. F. HAWKINS 457

The Interstimulus Interval and the Latency of the Conditioned Eyelid Response: C. A. BONEAU 464

Effects on Verbal Learning of Anxiety, Reassurance, and Meaningfulness of Material: I. G. SARASON 472

The Relationship of Stimulus Similarity and Number of Responses: J. RICHARDSON 478

Free Recall of Redundant Strings of Letters: G. A. MILLER 485

Training and Transfer as a Function of Component Interaction: G. E. BRIGGS AND L. K. WATERS 492

Acquisition of a Verbal Maze as a Function of Method of Correction and Number of Alternate Choices Per Unit: R. F. THOMPSON AND W. J. BROGDEN 501

Motivation Shift in a Complex Learning Task: D. BIRCH 507

The Acquisition of Imitative and Opposition Responses Under Two Conditions of Instruction-Induced Set: V. T. KANAREFF AND J. T. LANZETTA 516

Alterations in Critical Flicker Frequency as a Function of Age and Light:Dark Ratio: R. A. MCFARLAND, A. B. WARREN, AND C. KARIS 529

American Psychological Association

Vol. 56, No. 6 **December 1958**

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
SARAH J. DOYLE, Editorial Assistant

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, 53, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*. When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1958 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. and R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1958 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 56, NO. 6 DECEMBER, 1958

VEXIERVERSUCH: THE LOG RELATIONSHIP BETWEEN WORD-FREQUENCY AND RECOGNITION OBTAINED IN THE ABSENCE OF STIMULUS WORDS[1]

ISRAEL GOLDIAMOND AND WILLIAM F. HAWKINS[2]

Southern Illinois University

The relationship between word-frequency and recognition has recently received considerable experimental attention. Words more frequently used in the language have produced higher intelligibility (articulation) scores than those less frequently used (**2**), and similar relations hold for other studies of recognition of spoken words (**9, 12**), as well as for visual recognition of words flashed on a screen (e.g., **10, 16**). The extent to which the frequency-recognition relationship accounts for data from both modalities is the basis for considerable controversy in perceptual defense studies (**8**). This controversy relates to the more basic question of whether the data obtained are to be given a perceptual interpretation or a response interpretation. Stated otherwise, does frequency affect seeing or saying (cf. **13**)? This argument may have relevance for the validity of sensory inferences drawn from articulation and intelligibility scores, as well as from other recognition data.

The frequency with which a word appears in the language can only be estimated; word counts from selected publications (**6, 18**) provide samples which may not be representative of all *S*s (**4, 5**). In the investigation of the frequency-recognition relationship, nonsense syllables present an advantage in that their frequency is more amenable to experimental control, since they have less of a history of prior usage than other words. In a training session, for example, one syllable may be presented 25 times, another 10 times, and so on. Each syllable is then presented in a recognition session. Such experiments report frequency-recognition relationships of the type discussed (**1, 7, 11, 14, 16**).

This study reports such a nonsense syllable experiment. Following a training session of the type described, *S* was told that one of the training words would be flashed subliminally

[1] This research was made possible through grants from the National Science Foundation, to whom the author is greatly indebted for generous support.

[2] Now at the University of Virginia.

at regular intervals on the screen. He was to respond with a word whether he saw one or not; when he was correct, *E* would so inform him and would initiate a new series. This experiment, however, differs in that *all flashes were blanks*. If the previously reported frequency-recognition relationship is now obtained, this relationship cannot be ascribed to perception.

Recognition involves accuracy, and accuracy is usually regarded as the congruence between *S*'s responses and the stimulus sequence. Hence the perceptual inference. In actuality, however, *E* seldom co-monitors the signals with *S*; accuracy is usually "the congruence between *S*'s response and *E*'s score sheet" (**8**, p. 35).

If (*a*) words more frequently used in a training session (where the relation between reinforcement and use is constant) are also the more frequent responses in a recognition session, and if the training session words appear equally on *E*'s score sheet, then (*b*) the frequent words will score more hits than the infrequent words. The frequent words should also produce such congruences earlier in the series. If, as in the ascending Method of Limits, a hit terminates the series, and ascending stimulus energy values are assigned to successive presentations in the series, then (*c*) recognition thresholds of more frequent words should be lower than those for less frequent words.

This study has its origins in the Vexierversuche of classical psychophysics, where blanks were introduced as a check on *S*'s biases (**3**, p. 479). The training session in the current experiment would be considered as one of many possible ways to bias *S*'s responses, that is, to give responses unequal likelihood of appearing in the recognition response sample. Results consistent with the three predictions made have appeared in the recognition literature. In this experiment, the predictions are made without reference to a signal, and since none is presented, perception is precluded from the interpretation. The second and third predictions relate *E*'s *scoring procedures* and *score interpretations*, respectively, to the *response bias* of the first prediction, a function of training procedures familiar to students of learning.

Method

Subjects.—The *S*s were 25 male college students, volunteers from a sophomore lecture section of a physical education course.

Stimulus words.—Stimulus words were nonsense syllables of low association value drawn from a list compiled by Hilgard (see **17**). Ten words were divided into five pairs, and systematically varied so that each pair would be presented at each of five frequencies; 13 other words, introduced to lengthen the training series, were presented once and did not appear in the recognition series. The paired words were: MIV, WUX; TUD, ZOF; HIF, MAF; FEP, HAJ; VUK, VOF. The extraneous words were: BEJ, DAJ, FEH, JEX, KEJ, LIJ, PUY, TEV, TOV, VAB, VEF, ZID, ZIM. The frequencies used were 25, 10, 5, 2, 1. This resulted in 99 cards per *S*.

Training procedure.—Each of the words was typewritten in ¼-in. capital letters in the center of a 3 × 5-in. white card, with the 25-frequency words appearing on 25 cards, and so on. The deck was shuffled separately for each *S* and was placed face down between *E* and *S*. Every 8 sec., *E* exposed a card in view of *S*, who then read the letters and the name, thus: "F-E-P, fep." The *S*s were told that these were words of a foreign language, with *E* interested in their pronouncing each word as it appeared. These responses will be called *training responses*.

Interpolated activity.—Following the training period, *S* was given the January, 1958 issue of *Consumer's Report*, opened to an article on subliminal advertising, a half-page clipping from the *Southern Illinoisan*, a local daily, containing an interview with *E* on this subject, plus a smaller clipping from a news weekly, and was requested to read these. After 10 min., he was transferred to the perception booth.

Perception procedure.—The booth has controlled and monitored lighting, and *S* sits 10 ft. from a uniformly illuminated screen, in the center of which, at eye level, is an 11 × 11-in. opal glass screen for rear projection. Instead of white light, a gray mottled section from Rorschach Plate I was flashed at each presentation.[3] Interruption of the light was through a Wollensak solenoid-operated shutter connected to Hunter interval timers. Communication with *E* was through a high-fidelity two-way system.

The *S* was told that 2 sec. after the warning bell, one of the foreign words he had pronounced would be flashed into the window, so quickly that he might see only a flash; whether he saw a word or not, he was to respond with one, since the purpose of the experiment was to relate behavior to unseen or barely seen stimuli. Within three presentations, *S*s were responding with a word occasioned by the window lighting up. That a presentation was made was discernible; duration was .02 sec., with interval being 7 sec. All temporal relations and bell were automatic.

Score sheet and psychophysical method.—The method mimicked was the ascending Method of Limits; *S* was told that the same word would be presented until he got it right, that he would be informed when correct, and that another word (or the same word, by random selection) would then be shown. This procedure was used to avert perseverations; the data obtained can be analyzed as though a random presentation method were being mimicked. The *E* score sheet had 10 columns, each headed by a word. These were balanced so that each word would appear equally as the first or second, the third or fourth, and so on in the 10 series. When *S* said the word at the top of the column, *E* stated: "That was the correct word. We shall now try another series," and began to make entries in the next column. All words given were recorded in the appropriate column. A pilot study having produced no "correct" responses beyond Presentation 22, if the word at the head of the column were not given by Presentation 25, *E* said: "We shall now try another series," and began to make entries in the next column.

It will be noted that by this procedure, all *S*s did not give an equal number of responses, nor was there the same number of entries in each column, since giving the word at the head of the column terminated entries for that column, as in the studies using the ascending Method of Limits. In these studies, all entries preceding the correct one are called "prerecognition hypotheses." Responses given in the booth will be called *perceptual responses*.

Three *S*s failed to produce a single correct response, and a fourth produced only one. Their records contained runs of the same syllables. Since the same learning procedure was used for all *S*s who vary in learning in both training procedures and instructional periods, the performance of these *S*s was considered indicative of poor learning procedures, and they were replaced by four other *S*s prior to the analysis of the data.

Results

Perceptual-response frequency as a function of training-response frequency.—This relationship is presented in Fig. 1; the logarithmic relationship between the two frequencies usually found in recognition experiments of this type has been obtained here. The degree of association between

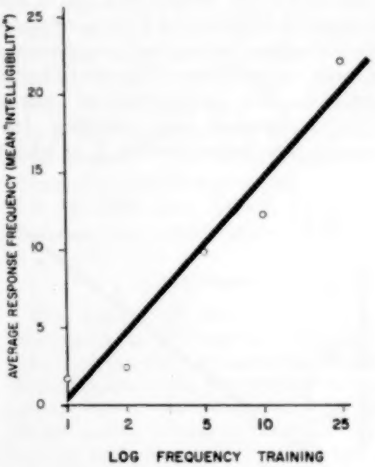


FIG. 1. Mean frequency of perceptual responses as a function of training frequency. If treated as ascending Method of Limits, mean frequency of "prerecognition hypotheses" as a function of training frequency. If treated as random series, mean "recognizability" ("intelligibility") as a function of training frequency

[3] This was used because it was available, and produced a random-appearing gray; the basic motivation being that "something" should be presented. This precaution turned out to be unnecessary; *S*s in a later experiment reacted as well to pure white light.

obtained order and predicted order, as measured by Kendall's Tau, is 1.00, which has a significance value of $P < .01$ for five ranks. For the training frequencies of the experiment, the average perceptual response frequency is: 25, 22.18; 10, 12.30; 5, 9.88; 2, 2.38; 1, .83 (two perceptual words, 1.04; 13 extraneous words, .80).

These results confirm the prediction made and suggest that had a random method been mimicked, instead of the ascending Method of Limits, *S*'s congruences would have tended to follow the distribution in Fig. 1, and "intelligibility" and "recognition" scores would have followed suit as an artifact of this relationship. If the data are interpreted as relating to the ascending method, Fig. 1 can be considered as depicting "prerecognition hypotheses."

Number of congruences.—When *S* gave the word at the head of a column, that series was terminated, as in the ascending Method of Limits recognition studies, where such congruence defines recognition. The relationship between the percentage of words thus recognized and training frequency is indicated in Fig. 2, in which the logarithmic relationship found in Fig. 1 is continued.

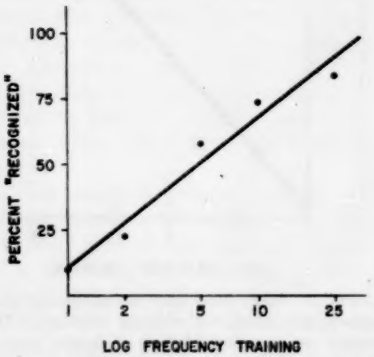


FIG. 2. Percentage of "presented" syllables "recognized" as a function of training frequency. Data treated as ascending Method of Limits.

TABLE 1

NUMBER OF RECOGNITIONS FOR EACH FREQUENCY AS A FUNCTION OF NUMBER OF RECOGNITIONS PER *S*

Words Recognized	*Ss*	Word-Frequency				
		25	10	5	2	1
7	5	10	10	8	4	3
6	5	8	10	7	4	1
5	4	7	8	3	2	0
4	5	8	6	6	0	0
3	6	9	3	5	0	1
Totals	25	42	37	29	10	5
Number unrecognized		8	13	21	40	45
Mean presentation at which recognized word was recognized		4.50	6.78	6.52	6.40	10.80

No *S* recognized all words. Table 1 relates the total number of words recognized by each *S* to the number of words recognized by *S* for each frequency. If *S* recognized all 10 words, the recognitions would be evenly distributed among all five frequencies. If he recognized only one, this should tend to be a 25-training frequency word, with a curve in between gradually moving away from the 25-frequency words. The table indicates that recognitions generally distributed themselves in accordance with this expectation.

Average threshold.—With regard to the relationship between thresholds and frequency, to what extent are the results of this experiment similar to results obtained where signals are actually presented? Assuming that signals were presented here would require that stimulus energy be increased with each presentation by some physical value we shall call an erg. It will be noted from Table 1 that many words required energy levels higher than 25 ergs for recognition—they were not recognized by

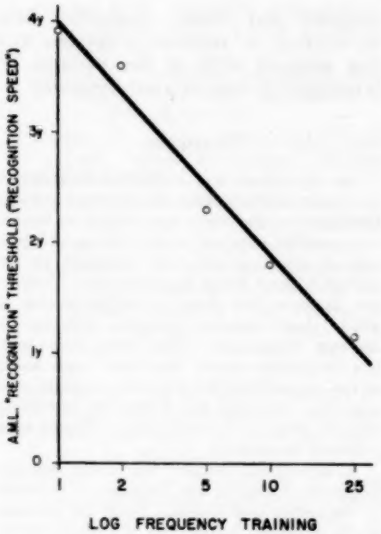


FIG. 3. "Threshold energy" required to "recognize" words as a function of training frequency. Data treated as ascending Method of Limits.

Presentation 25. If we assume that *all* words, regardless of frequency, require an average of 26 ergs for recognition, we can compute an average threshold for each word, based upon number of words unrecognized at Presentation 25 (given values of 26), plus number of words unrecognized at preceding presentations, but previously congruent through the interaction of positive response bias with score sheet entries (given values of the presentation of congruence). Such average thresholds are presented in Fig. 3. The ordinate is expressed in terms of relative stimulus energy; *y* can stand for ergs, milliseconds, millivolts, and the like. The inverse logarithmic relation between stimulus energy and frequency reported in the literature, and often considered indicative of perceptual effects, has been obtained.

Other results.—The syllables were selected from a list of words of equal association value; that their values were not equal for the population in this study is suggested by marked differences in the frequency of perceptual responses given. Each of the recognition words was presented an average of 8.3 times in the training session; the number of perceptual responses for each was: TUD, 394; HAJ, 382; MIV, 294; WUX, 270; VOF, 230; FEP, 200; MAF, 192; HIF, 153; VUK, 143; ZOJ, 127. The 13 nonrecognition words were presented once in the training session; their perceptual response frequencies range from TEV, 103; PUY, 42; to KEJ, ZUM, LIJ, each 0.

A total of 1259 neologisms were given by all *Ss*, constituting .31 of all responses. All but 156 of these shared two letters with the training words. Consideration of these as partial responses (or partial "recognitions") did not alter the logarithmic functions when added to the data. No consistency could be found in their use. Words scoring high on partial response scored both high and low on full response; similar results obtained for other scores. The analysis was complicated by duplications (for example, T-V in TEV and TOV), for which arbitrary corrections were made.

DISCUSSION

The results of this study can be interpreted as challenging a perceptual interpretation of the relationship between word-frequency and recognition-intelligibility, where word-frequency can be placed under laboratory control. Perception was not involved in this study, yet the logarithmic recognition-frequency curves were obtained. If we pretend, as in Fig. 3, that syllables were flashed on the screen, the data make sense when we assume that all syllables are equally perceptible. Stated otherwise, we as-

sume that frequency as a variable does not affect perceptibility. If, continuing with the signal pretense, we interpret the data to indicate that frequency does affect perception—that is, the results indicate differences in perceptibility—we find ourselves in the position of stating that the same data can be adequately explained by having frequency both affect perception and not affect it. On the other hand, stating that frequency does not affect perception, but does affect response bias, eliminates the contradiction as well as explains the data. The lowered stimulus energy at which recognition of high-frequency words occurs has been interpreted to mean that "the more frequently a word is used . . . the more readily it can be seen" (15). This interpretation is open to question since the inverse energy-frequency relation from which it is drawn may simply be an artifact of coupling an ascending energy series to sequential progression of a series.

The results of this study may also be relevant for interpretation of articulation tests. Using *S* as referent, congruence between responses and *E*'s score sheet is used to define sensitivity; using the word as referent, such congruence is used to define intelligibility. In the one case, inferences are often made about the person; in the other, about the perceptibility of the stimulus.

Both identification and recognition involve an accuracy indicator which may be heavily influenced by congruences related to response biases and irrelevant to the perceptual issues under study. Although this experiment deals with nonsense syllables, the frequency of which is experimentally varied, the author would consider the data as suggesting, at least, that similar considerations may hold for studies relating other types of frequency to recognition thresholds; implications of the consideration of the role of response bias are drawn for other experiments as well in a recent methodological analysis of perceptual experiments (8). Most recently, Howes (9) has interpreted parallels between auditory and visual recognition data as relating to response variables, and has assigned 69% of the variance in intelligibility data to word-frequency.

Summary

An experiment was conducted to ascertain the extent to which previously reported relations between word-frequency and recognition thresholds could be obtained in the absence of a perceptual stimulus, and could therefore be explained without resort to perception. Twenty-five *S*s were first given a training session in which they repeated nonsense syllables at different frequencies. They were then told, in a recognition session, that these words would be flashed subliminally; they were to guess until accurate. Accuracy was defined as responding with the word on *E*'s score sheet. Blanks were presented throughout.

A logarithmic relation was found between frequency of training and frequency of response in the recognition session. Since the training words appeared equally on *E*'s score sheet, the more frequent responses led to more hits, this accuracy-frequency relationship also being logarithmic. Since an ascending Method of Limits was mimicked, with accuracy terminating a series, the more frequent words also scored hits earlier, producing lower thresholds.

The similarity between these results, which cannot be ascribed to perception, and the data obtained from similar experiments where a stimulus has been presented, was interpreted as challenging a perceptual interpretation of the word-frequency-recognition relationship, where similar procedures are utilized. Implications of these results for recognition, identification, intelligibility, and articulation studies were discussed, with the suggestion offered that the existence of a contribution to the variance from response bias necessitates caution in inferring sensitivity or intelligibility (perceptibility) from such data.

References

1. Baker, K. E., & Feldman, H. Threshold-luminance for recognition in relation to frequency of prior exposure. *Amer. J. Psychol.*, 1956, **69**, 278–280.
2. Black, J. W. Accompaniments of word intelligibility. *J. Speech Dis.*, 1952, **17**, 409–418.
3. Boring, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century-Crofts, 1942.

4. Daston, P. G. Perception of idiosyncratically familiar words. *Percept. mot. Skills*, 1957, **7**, 3–6.
5. Davids, A. Personality dispositions, word frequency, and word association. *J. Pers.*, 1956, **24**, 328–338.
6. Dewey, G. *Relativ frequency of English speech sounds*. Cambridge: Harvard Univer., 1923.
7. Forrest, D. W. Auditory familiarity as a determinant of visual threshold. *Amer. J. Psychol.*, 1957, **70**, 634–636.
8. Goldiamond, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–411.
9. Howes, D. On the relation between the intelligibility and frequency of occurrence of English words. *J. acoust. Soc. Amer.*, 1957, **29**, 296–305.
10. Howes, D. H., & Solomon, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401–410.
11. King-Ellison, P., & Jenkins, J. J. The durational threshold of visual recognition as a function of word-frequency. *Amer. J. Psychol.*, 1954, **67**, 700–703.
12. Mason, H. M., & Garrison, B. K. Intelligibility of spoken messages: Liked and disliked. *J. abnorm. soc. Psychol.*, 1951, **46**, 100–103.
13. Neisser, U. An experimental distinction between perceptual process and verbal response. *J. exp. Psychol.*, 1954, **47**, 399–402.
14. Postman, L., & Rosenzweig, M. R. Practice and transfer in visual and auditory recognition of verbal stimuli. *Amer. J. Psychol.*, 1956, **69**, 209–226.
15. Rosenzweig, M. R., & Postman, L. Frequency of usage and the perception of words. *Science*, 1958, **127**, 263–266.
16. Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–201.
17. Stevens, S. S. (Ed.) *Handbook of experimental psychology*. New York: Wiley, 1951.
18. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Columbia Univer., 1944.

(Received March 14, 1958)

Journal of Experimental Psychology
Vol. 56, No. 6, 1958

THE INTERSTIMULUS INTERVAL AND THE LATENCY OF THE CONDITIONED EYELID RESPONSE [1]

C. ALAN BONEAU
Duke University

The effect upon eyelid conditioning of the time interval between the CS and the UCS has been studied in detail, with a rather precise specification of the shape of the function emerging from complementary studies by Reynolds (**12**) and Kimble (**6**). Beyond this, recent studies, using the technique of changing from one interstimulus interval to another during conditioning, have produced results which show that the level of conditioning can be altered by such manipulation of the interval. McAllister (**10**), for example, began conditioning with an optimal interval and then shifted to interstimulus intervals at which no conditioning usually occurs. For a group shifted from a .5-sec. to a 2.5-sec. interstimulus interval, the result was a gradual extinction of the CR, a marked contrast to the abrupt decline exhibited by a group extinguished in the usual manner with CS alone. Boneau, King, and Kimble (**2**) have reported a similar experiment in which the shift was to an interval permitting acquisition of a CR. The results of this experiment were similar to those of McAllister. There was also some evidence that responses appropriate to the different interstimulus intervals behave independently of one another after the shift.

[1] Paper based upon a dissertation submitted to Duke University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Research was carried out while the author was research assistant on Grant NSF-G-2147. The author is indebted to Gregory A. Kimble, as principal investigator, for his part in directing the project and for his help in preparing the report.

All of these studies have been primarily concerned with the amount of conditioning defined in terms of percentage of CR's in successive blocks of trials. The possibility that the interstimulus interval influences other measures such as latency has largely been ignored. The experiment reported here studied the effects of shifting from an optimal interstimulus interval to two different nonoptimal ones and made a detailed analysis of the latency of CR's associated with different interstimulus intervals.

METHOD

Apparatus.—The eyelid-conditioning apparatus previously employed at Duke University (**8**) was used in this experiment. This apparatus photoelectrically responded to the movement of the eyelid. A small aluminum-foil eyelash attached to *S*'s right eyelid interrupted a beam of light directed at a photocell. Movement of the eyelash out of the beam (approximately 1/8 in. downward excursion at a distance of 2 in. from the eyelid) was recorded as a blink. The *S* sat in a sound-deadened, lighted cubicle with his head resting in a chin-and-head rest. Modifications of the apparatus for the present experiment included the addition of an Esterline-Angus Operations Recorder which recorded onset and termination of both the CS and UCS as well as the occurrence and termination of blinks. In addition to this graphical record, a Standard Electric .01-sec. timer was utilized to get direct measurements of the latency of the blink recorded from time of onset of CS.

The CS was a small light ¼ in. in diameter at a distance of 20 in. from *S*. It remained on until the termination of the UCS. The UCS was an airpuff, nominally .05 sec. in duration at a pressure equivalent to 50 mm. of mercury. It was released by a solenoid valve and delivered from a distance of approximately ½ in. to the corner of *S*'s right eye through a nozzle with a 3/64-in. aperture. Intertrial intervals, programmed by a motor-operated disc, repeated the sequence 12, 10, 16, and 22 sec.

Subjects.—The *Ss* were undergraduate students at Duke University participating to fulfill a course requirement. Of the 87 *Ss*, five were rejected because of a high blink rate on a pretest (see Procedure), five were classified as voluntary responders on the basis of the criterion proposed by Spence and Taylor (**14**) and were discarded. Thirteen others were not used because they failed to give UCR's on five or more trials during the conditioning procedure, presumably because of the weak airpuff.

Procedure.—Of the remaining 64 *Ss*, 34 received 50 trials with an interstimulus interval of .5 sec. and were then divided into two groups equated in terms of mean number of CR's in this phase of the experiment. One of these groups, the *.5-1.0 Shift Group*, was then switched to a 1.0-sec. interstimulus interval for 100 further trials. The other group, the *.5-1.5 Shift Group*, received 100 further trials with a 1.5-sec. interval. Two control groups, each consisting of 15 *Ss*, were designated the *1.0 Control Group* and the *1.5 Control Group*. These groups received 150 trials at their appropriate interval. A prearranged schedule insured randomness of assignment of *Ss* to the shifted group and to the two control groups.

Instructions were those used previously (**8**) with the addition of a statement describing the pretest procedure. While instructions were recited to *S*, he was at the appropriate times moved into the cubical, fitted with a false eyelash, and adjusted into the apparatus. The *S* then received four airpuffs alone spaced at the usual intertrial intervals. This was followed by a pretest series of ten presentations of the CS alone at a duration of .5 sec. The *Ss* were immediately discarded if on the pretest they gave more than four responses to the CS with a latency of .5 sec. or less or gave more than six responses with a latency of less than 1.0 sec. Following the pretest, *S* received a 1-min. rest. The *S* then received his 150 trials with a 1-min. rest following Trials 40, 80, and 120. No ready signal was used. The total time elapsing from beginning of instructions to completion of the 150 trials was approximately 50 min.

Results

Definition of the CR.—The CR's in eyelid conditioning have generally been considered to be responses which anticipate the onset of the UCS but which are not UCR's to the CS. If test trials are used, CR's are those responses which anticipate the onset of the UCS or occur in its absence. Since the latency characteristics of responses to different intervals are different and since the present experimental procedure involves a shift from an optimum interstimulus interval to longer ones, the problem of the definition of a CR for the specific intervals employed had to be faced at the outset of the analysis of results. In dealing with this problem the assumption was made that CR's will tend to increase with increasing numbers of trials. Based upon this assumption, a CR was defined as a response in any range of response latencies which shows a consistent increase in level of responding as conditioning proceeds. Operationally, the distribution of responses by latencies for an early stage of conditioning is compared with that for a later stage or stages. That part of the distribution which increases from one stage to the next constitutes a range of response latencies which is used to define the CR's for the interstimulus interval involved. Such a procedure automatically compensates for the larger number of random blinks expected in the interval between CS and UCS as its length increases by tending to eliminate them from consideration as CR's.

Because of the importance of the latency distributions, they are presented in Fig. 1, 2, and 3 for the .5-, the 1.0-, and the 1.5-sec. interstimulus intervals, respectively. For all interstimulus intervals, the responses are presented as a set of five separate distributions representing successive fifths of the conditioning trials. It should be noted that the set of distributions for the .5-sec. interval is based only upon the first 50 trials of the 34 *Ss* who received their original conditioning at that interval before being shifted. The other distributions are based upon the full 150

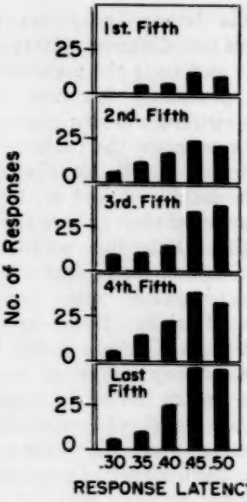


FIG. 1. Latency distributions of responses to 5-sec. interstimulus interval.

trials for each *S* in the appropriate control group. In these figures, the columns represent the number of responses for latency ranges in .05-sec. steps. The designating number below each bar indicates the maximum response latency (measured in .01 sec.) in that column. Distributions are truncated at .25 sec. since a negligible amount of responding occurred before that point.

Characteristic of the course of responding as shown by the sets of distributions is a progressive increase in the number of responses in the region immediately adjacent to the point of application of the UCS. There can also be noted a tendency for the responses with shorter latencies to show an initial increase followed by a decrease in level of responding for all three interstimulus intervals. This is paralleled by a shift of the median response latency as indicated by the arrows. Another expression of this tendency is seen, at least for

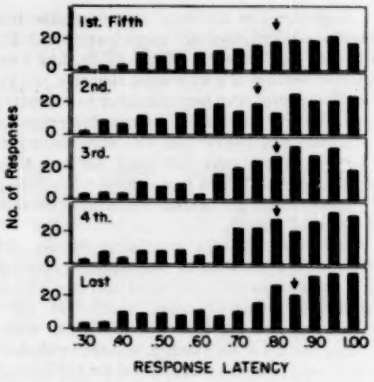


FIG. 2. Latency distributions of responses to 1.0-sec. interstimulus interval.

the .5- and 1.0-sec. interstimulus intervals, to be a successive narrowing of the range of response latencies which shows an increase in level of responding when compared with the first fifth of trials. This suggests that the anticipatory blinks tend more and more closely to approximate the onset of the UCS. Data of a different kind illuminate this process. The records from the operations recorder permitted an examination of the temporal relationship between blink and puff. Specifically, it was possible to determine whether or not the blink preceded the puff and whether its duration was long enough to overlap the puff. Typically, the

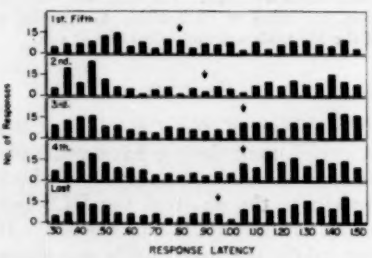


FIG. 3. Latency distributions of responses to 1.5-sec. interstimulus interval.

records revealed that blinks lasted approximately .2 sec., that a sizeable number did *not* overlap the puff and thus did not successfully avoid it. Analysis of the records for the 1.0 Control Group showed a progressive increase in the percentage of avoiding CR's from 70% for the first fifth to 90% for the last fifth of conditioning. Of the 12 *S*s showing a change in percentage of avoiding CR's, 11 increased, an event whose occurrence by chance has a probability of less than .01.

Thus there is evidence that, as conditioning proceeds, there is an increase in the percentage of CR's, and also an increase in the precision of the timing of the blink so that it tends more and more to overlap the puff, presumably vitiating its effects. Such a finding complicates the attempt to establish a simple definition of the CR since the region of increased responding progressively narrows with increasing numbers of trials. Rather than use a sliding criterion, the following ranges were rather arbitrarily decided upon as definitions of CR's appropriate to their respective interstimulus intervals: .31 to .50 sec. for the .5-sec. interstimulus interval, .66 to 1.00 sec. for the 1.0-sec. interval, and 1.01 to 1.50 sec. for the 1.5-sec. interval. As reference to the latency distributions will show, these ranges fairly adequately satisfy the principle which designates as CR's only those responses which show a *consistent* increase in level of responding with practice. Such a definition is for the 1.5-sec. interval least satisfactory. However, the set of distributions for that interval reflect a large number of double blinks, the first occurring at approximately .4 sec. and the second about 1.0 sec. later, which 6 of the 15 *S*s produced. Since the significance of the first blink was rather obscure, it was decided to ignore it at least for the purpose of establishing a definition of CR comparable to those for the other interstimulus intervals. It is interesting to note that, due to the double blinks, the level of responding for the 1.5 Control Group was highest in the region just preceding .5 sec. Figure 4 presents the total responses summed over all 150 trials for the 1.5 Control Group, grouped by intervals of 1/10 sec. The marked similarity of the first half of this curve to idealizations of the curve relating amount of conditioning to interstimulus interval, as in Kimble (**7**, p. 212), is provocative.[2]

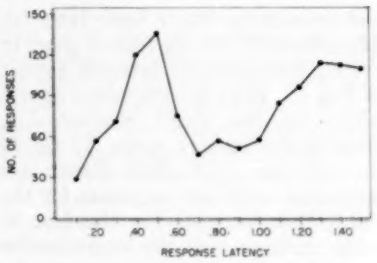


FIG. 4. Distribution of response latencies of 1.5 Control Group for all trials.

Learning curves.—Conditioning functions plotting percentage of CR's as defined above on successive blocks of trials are similar to those usually obtained (**6, 12**) with relative performance best for the .5-sec. interval

[2] Hull (**5**, p. 106) predicted a response curve similar to that of Fig. 4. He assumed the CS in actuality to be the internal stimulus trace resulting from the application of the experimental CS. Since the intensity of the trace presumably rises for the first half-second, then declines, there should occur some responses early in the interstimulus interval when the trace has about the same intensity as at the point where the UCS is applied. Then should occur a generalization decrement as the trace increases in intensity before declining to its value at the point of application of UCS.

and poorest for the 1.5-sec. interval. The effects of shifting from a short to a long interstimulus interval appear in Fig. 5. Here the responses appropriate to the short interval—i.e., lying in the latency range .31 to .50 sec.—of the .5-1.0 Shift Group are compared with like responses for the .5-1.5 Shift Group. For the first 50 trials during which the interstimulus interval is .5 sec. for both groups, the percentage of CR's increased for both groups. After the shift to a longer interstimulus interval, however, these short interval responses gradually approached a level of less than 10%. There is a suggestion that these responses extinguished more rapidly for the group for which the shift was greatest, but the difference was not significant. By the Kolmogorov-Smirnov test, used because of the marked non-normality of the extinction data, the difference was significant at approximately the 20% level.

After the shift from a short to a long interstimulus interval, responses appropriate to the new interstimulus interval exhibited an increasing percentage by blocks of 10 trials. The acquisition curve for these responses following the shift of interval was compared with the acquisition curve for the appropriate control group to determine whether or not the previous conditioning at the shorter interval had an effect on the acquisition of

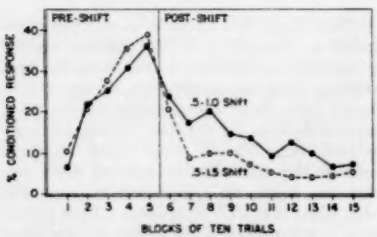


FIG. 5. Comparison of response curves for .5–1.0 Shift Group and .5–1.5 Shift Group.

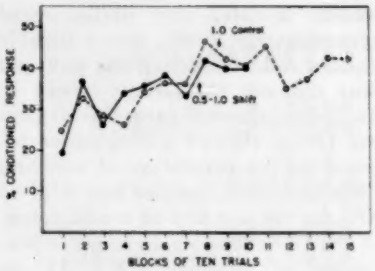


FIG. 6. Comparison of responses with latency .66 to 1.00 sec. for .5-1.0 Shift Group and 1.0 Control Group.

the long interval responses. Figure 6 makes this comparison between the 10 blocks of 10 trials following the shift for the .5-1.0 Shift Group and the acquisition curve for all responses for the 1.0 Control Group. There can be no question of the lack of a significant difference between the two curves. A similar but not quite so striking result (not presented here) holds for the groups conditioning with an interstimulus interval of 1.5 sec. Apparently, the performance during conditioning for the longer interstimulus intervals increases at the same rate whether or not *Ss* had been conditioned previously with a shorter interstimulus interval.

This same finding by Boneau, King, and Kimble (**2**) had led to a consideration of the possibility that the distribution of response latencies simply changed in temporal position by sliding through the region intermediate between the two which define CR's. If this were so, the intermediate region should exhibit a greater level of responding for the shift group following the shift than is exhibited by the control group. Such was not the case; nor is it in the present experiment. Figure 7 presents the level of responding for all blocks of 10 trials of the 1.0 Control Group in the inter-

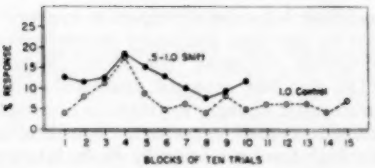


FIG. 7. Comparison of responses with latency .51 to .65 sec. for .5–1.0 Shift Group and 1.0 Control Group.

mediate region of latencies, .51 to .65 sec., as well as the corresponding responses for the 10 blocks of trials of the .5-1.0 Shift Group which occurred after the shift. The similarity between the two curves is marked even to the slight rise in response level for both groups during the fourth block of ten trials after the beginning of conditioning at this interstimulus interval. For both groups, the peak can be attributed to the previously observed increased level of responding at the shorter latencies early in conditioning. As before, the results for the 1.0 groups were paralleled by the results for the 1.5 groups, which are not presented here.

DISCUSSION

The results of shifting from an optimal interstimulus interval to a longer, nonoptimal one may be summarized briefly. First, as in McAllister's study (**10**), the responses appropriate to the shorter interval extinguish gradually. Secondly, the responses appropriate to the longer interstimulus interval, simultaneously with the above, describe an acquisition curve indistinguishable from that of a control group conditioning only at the longer interval. Apparently conditioning at the shorter interval has no effect on the acquisition of the response appropriate to the longer interval. Thirdly, in the interval of latencies between those intervals which define CR's, the level of responding seems to be the same as that for the appropriate control group after a similar amount of conditioning at the longer interval. This latter fact is evidence that the distribution of response latencies does not shift in temporal locus in a continuous fashion from the region appropriate to the first interstimulus interval to that appropriate to the second interval. The responses appropriate to the two interstimulus intervals thus seem to demonstrate independence, one extinguishing at the same time the other shows acquisition, with no evidence of interaction between the two processes. These results constitute in detail a replication of the results of the Boneau, King, and Kimble study.

From the above results, the prediction may be made that the result of a shift from one interstimulus interval to a longer one will be, if all anticipatory responses are considered, simply a gradual change to a level of responding which would have been obtained had the conditioning been only at the longer interstimulus interval. This is so since the short interval responses will extinguish to a level of random responding, the long interval responses will approach their appropriate asymptote, and the level of all other responses will be the same under both conditions.

The additional finding of the present experiment that the region of increased responding, which is used to define CR's, narrows with increased training raises an important issue. Such an increase in latency as a result of training is in disagreement with certain available evidence. In the well-known studies of Pavlov (**11**, pp. 91–102), the salivary response for intervals of a few minutes or less typically decreases in latency so as to follow the CS immediately. In another study to which frequent reference is made, Campbell and Hilgard (**3**) report a fairly strong inverse relationship ($r = -.54$) between latency and frequency of CR in the eyelid conditioning situation, seemingly the opposite of the present results. The latter discrepancy may be accounted for by a consideration of the conditions under which the Campbell and Hilgard experi-

ment was conducted. The photographic method of recording utilized by them necessitated conditions conducive to dark adaptation of *S*. As Grant and Norris (4) have shown, dark-adapted *S*s tend to respond to light with a blink ("beta" response) having a latency of .12 to .24 sec. The fact that the bulk of responses reported by Campbell and Hilgard fell in this range would suggest that the decrease in latency of responding reported by them was due to increasing numbers of beta responses rather than a decrease in the latency of the true CR.

In considering the Pavlov results, it must be remembered that the salivary response of the dog is very dissimilar to the human eyelid response, and the possibility exists that the two responses may function quite differently. A striking feature of the eyelid response, for example, is the adaptiveness of the change of precision of the blink so that it tends more and more to avoid the puff. Once the CR appears, it behaves as an instrumental response in that its characteristics can be altered by the reinforcing contingencies in the situation since analysis of the nonavoiding responses has demonstrated that the change in latency is indeed related to the reinforcement properties. This assumes that avoidance of the puff is a reinforcing state of affairs, and that responses with latencies such that the puff is avoided tend to occur predominantly as a result of differential reinforcement. In much the same way, bar presses of a specified pressure tend to occur when reinforcement is made contingent upon a response of that specific pressure (1). The micromolar approach of Logan (9) would seem to find a meaningful application here. If blinks of different latency are considered to be different responses, reinforcement of blinks having latencies such that the puff is avoided would thereby increase in response strength, while nonavoiding blinks would tend to extinguish. The consideration of responses of different latencies as being different responses is rendered more tenable by the observation of the seemingly independent behavior of responses appropriate to the long and short interstimulus intervals following the shift. Spence (13) also has reported results showing systematic changes in latencies of conditioned eyelid responses. Specifically, for high-level conditioning *S*s the latency decreased, while for low-level conditioners an increase in latency resulted as conditioning progressed. While no explanation for these divergent results was presented, Spence suggested that the latency of the response may be what is learned. The present results support this contention.

If the preceding analysis is correct, the latency characteristics of the classically conditioned eyelid response may be modified by reinforcing contingencies inherent in the situation so that the response tends to maximize the reinforcing potential present. The Pavlovian results also are amenable to such a treatment. It is only necessary to make the reasonable assumption that saliva in the mouth of the dog at the same time food powder or acid is applied to the tongue is a reinforcing state of affairs. If this be so, then, up to a point, more saliva, the result of a CR with a shorter latency, should be even more reinforcing. The observed decrease in latency would easily follow from the principle of maximization of reinforcement.

A further interesting difference between eyelid and salivary conditioning can be predicted from such an analysis. We assume that the simultaneity of CR and UCS is essential for strengthening response tendencies beyond an initial level. In temporal conditioning the simultaneous occurrence of two extremely short events such as blink and puff should approach zero probability with increasing intertrial intervals. Precision of timing seems much less important in the case of salivation and food since salivation is a cumulative process. As a result, we would expect that strengthening of the temporal conditioned response would be relatively easy for salivation and virtually impossible for the eyeblink.

Summary

Three groups were given eyelid conditioning trials with interstimulus intervals of .5, 1.0, or 1.5 sec. The group conditioning at the .5-sec. interval was subdivided into two groups, each of which was further conditioned at one of the other intervals. Ranges of latencies defining CR's for each of the interstimulus intervals were established by considering the ranges of response latencies which showed a consistent increase in level of responding with practice. It was discovered that the latency of the response tended to increase with conditioning and that the percentage of blinks which succeeded in avoiding the puff also increased. This fact was discussed with reference to the usual assumption and evidence that latency decreases in the course of conditioning and an attempted reconciliation was made in the empirical conflict. Using the definitions of CR's from the preceding analysis, learning curves were constructed which showed that, after the shift from short to long interstimulus interval, the responses appropriate to the short interval showed a gradually diminishing level while, at the same time, the level of responding of responses appropriate to the long interval increased at the same rate as for the groups having no conditioning experience.

References

1. Arnold, W. J. An exploratory investigation of primary response generalization. *J. comp. Psychol.*, 1945, **38,** 87–102.
2. Boneau, C. A., King, R. A., & Kimble, G. A. The effect of changing the interstimulus interval on the acquisition of the conditioned eyelid response. *Amer. Psychologist*, 1956, **11,** 393.
3. Campbell, A. A., & Hilgard, E. R. Individual differences in ease of conditioning. *J. exp. Psychol.*, 1936, **19,** 561–571.
4. Grant, D. A., & Norris, E. B. Dark adaptation as a factor in the sensitization of the beta response of the eyelid to light. *J. exp. Psychol.*, 1946, **36,** 390–397.
5. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
6. Kimble, G. A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, **37,** 1–15.
7. Kimble, G. A. *Principles of general psychology.* New York: Ronald Press, 1957.
8. Kimble, G. A., Mann, L. I., & Dufort, R. H. Classical and instrumental eyelid conditioning. *J. exp. Psychol.*, 1955, **49,** 407–417
9. Logan, F. A. A micromolar approach to behavior theory. *Psychol. Rev.*, 1956, **63,** 63–73.
10. McAllister, W. R. Eyelid conditioning as a function of the CS-US interval. *J. exp. Psychol.*, 1953, **45,** 417–422.
11. Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) London: Oxford Univer. Press, 1927.
12. Reynolds, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.*, 1945, **35,** 15–30.
13. Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.
14. Spence, K. W., & Taylor, J. Anxiety and the strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42,** 183–188.

(Received January 16, 1958)

Journal of Experimental Psychology
Vol. 56, No. 6, 1958

EFFECTS ON VERBAL LEARNING OF ANXIETY, REASSURANCE, AND MEANINGFULNESS OF MATERIAL [1]

IRWIN G. SARASON

University of Washington

Several studies in recent years have contributed to our knowledge of the effects of anxiety as measured by questionnaires on the performance of *Ss* in stressful and nonstressful (neutral) situations (e.g., **2, 4, 10**). The following are a few of the generalizations one might make on the basis of some of this experimental evidence.

Under conditions of personal threat induced by experimental instructions (e.g., motivational instructions or failure reports) high-anxious *Ss* tend to perform at a lower level than do low-anxious *Ss* on tasks such as those often used in verbal learning experiments (**4, 5, 10, 11, 12**). Under neutral, nonthreat conditions (to the extent that such conditions can be experimentally created) high-anxious *Ss* appear to perform on at least as high a level as do low-anxious *Ss* on similar tasks (**4, 5, 10, 11, 12**). There is a suggestion that different anxiety scales, e.g., test and general anxiety questionnaires, may lead to different kinds of relationships in certain kinds of situations (**8, 13**). It may well be that the more specific the anxiety scale is to situations which are similar to the experimental situation, the more will the measure of anxiety be related to performance.

The evidence for these generalizations also provides a basis for an initial attempt to conceptualize the variable or variables being tapped by anxiety scales. High-anxious *Ss* may differ from other *Ss* in the extent to which their performance is disrupted by task-irrelevant responses when they perform under stress (**2, 5, 16**). It may be conjectured that there are at least two possible classes of interfering responses: (*a*) self-verbalization, e.g., "This test is too hard for me," "I am stupid," and (*b*) autonomic anxiety responses, e.g., accelerated heart rate, sweating, etc. Since these hypothesized interfering responses are assumed to be elicited to a much greater degree in high-anxious *Ss* than in low-anxious *Ss* under conditions of personal threat, and if they are indeed "interfering" responses, the relatively greater disruption in the performance of high-anxious *Ss* after personal threat would be expected. Under nonthreatening conditions the arousal of interfering responses would not necessarily be expected either in high- or low-anxious *Ss*.

Present knowledge of the antecedent conditions of such interfering responses is very meager. If, however, it is assumed that the interfering responses aroused under certain motivation conditions are learned responses, it seems fruitful to explore the conditions which would lead to the extinction of those interfering responses. This was the principal purpose of the present experiment.

If high- and low-anxious *Ss* are placed in a learning situation and

[1] This research was supported in part by grants from the Agnes H. Anderson Research Fund and the Graduate School Research Fund of the University of Washington. The writer is indebted to Robert E. Farley for his assistance in carrying out this experiment.

are given only the instructions necessary to communicate how they are to carry out the task, important differences between high- and low-anxious *Ss* would not necessarily be expected. However, the high-anxious *Ss* may, even so, interpret the experimental situation as personally evaluative. Therefore, it can be inferred that even under neutral conditions high-anxious *Ss* emit responses which could compete with task-appropriate responses. Available data indicate that under neutral conditions high-anxious *Ss* appear to do as well as low-anxious *Ss* on many intellectual tasks (**5, 10, 11, 12**). Therefore if such interfering responses were reduced in number or strength prior to performance, high-anxious *Ss* should perform at a *higher* level than would low-anxious *Ss*. The present experiment attempts to reduce the effects of interfering responses by the use of "reassurance" instructions.

Another important variable in evaluating the effects of anxiety and experimental instructions on performance is the nature of the task (**3, 4, 6**). A very difficult task, for example, may lead to a greater decrement in performance for high-anxious *Ss* than for low-anxious *Ss*, not because of the task's difficulty per se but because a more difficult task may arouse much more anxiety (i.e., more interfering responses) than would a simple task for high-anxious *Ss* (**9**). Also, some evidence suggests a possible interaction between anxiety and difficulty of task as defined, for example, by the strength of competing responses in a list of nonsense syllables (**3, 4, 6**). As a consequence of these considerations, two levels of difficulty of task were employed in the present experiment. Thus, there were three variables in our experiment: (*a*) anxiety, high and low scorers; (*b*) instructions, reassurance and neutral; and (*c*) task difficulty, two lists of words differing in their meaningfulness.

Two instruments were used to assess anxiety. One instrument was a Test Anxiety Scale (TAS) developed by the writer (**14**); the other instrument was Bendig's short form of the Taylor Manifest Anxiety Scale (MAS) (**1**). These two scales were used to assess the importance of the specificity of anxiety scales mentioned earlier. If the anxiety scale most closely related to the kinds of anxieties which might be aroused in a given situation leads to the most meaningful results, then the present results would be expected to more closely adhere to the stated expectations when *Ss* were categorized in terms of TAS than when categorized in terms of the MAS which contains no items specifically concerned with reactions to test situations. For 67 college women and 309 college men, Pearson product-moment correlations between TAS and MAS of .49 and .41, respectively, have been reported (**15**).

Method

Subjects.—The *Ss* were 76 students taking an introductory psychology course at the University of Washington. For the TAS scores the mean, median, and *SD* of the obtained distribution were 5.45, 5.61, and 3.23, respectively. The range was 0–16. The *Ss* with TAS scores of 6 and over were assigned to high-TAS groups, and *Ss* with scores below 6 were assigned to low-TAS groups. Since there were more *Ss* in the groups receiving standard instructions than those receiving reassurance instructions, it became necessary to randomly discard 12 *Ss* from the former groups in order to have equal *Ns* of 8 *Ss* per treatment group. This was done in order to carry out the analyses of variance involving repeated measurements to be described below.

For the same reason, 12 *Ss* were randomly discarded for the MAS analysis of variance. The mean, median, and *SD* of the MAS score distribution were 4.75, 4.47, and 3.65, respec-

TABLE 1

LISTS OF DISSYLLABLES DIFFERING IN MEANINGFULNESS AS DEFINED BY NOBLE'S *m* VALUES

Low Meaningfulness		High Meaningfulness	
.99 GOJEY	1.28 NARES	8.98 MONEY	6.24 INCOME
1.04 NEGLAN	1.28 ZUMAP	8.33 DINNER	6.02 HUNGER
1.79 BRUGEN	1.34 NOSTAW	7.95 OFFICE	5.98 REGION
1.24 XYLEM	1.30 POLEF	5.47 KEEPER	5.61 MALLET
1.24 TAROP	1.39 BODKIN	7.70 JELLY	5.94 LEADER
1.74 DAVIT	1.26 LATUK	7.39 INSECT	5.52 KENNEL
1.22 BALAP	1.13 BYSSUS	7.12 ZEBRA	8.12 WAGON
1.73 MATRIX	1.55 KUPOD	7.58 JEWEL	5.13 EFFORT
1.27 GOKEM	Mean = 1.34	6.57 UNCLE	Mean = 6.80

tively. The range was 0–16. The *Ss* assigned to high-MAS groups had scores of 5 and greater, while low-MAS *Ss* had scores of 4 and below.

Test materials.—The Test Anxiety Scale (TAS) consists of 17 True-False items dealing with *S*'s reactions before, during, and after taking examinations (**14**). The present TAS represents a transformation of the graphic rating scale Test Anxiety Questionnaire (**5**) into true-false terms.

The Taylor MAS measure used was Bendig's 20-item short form (**1**). Bendig has presented evidence which indicates that his 20-item MAS scale possesses slightly less reliability than does the 50-item MAS but probably contains the 20 most valid items of the 50 Taylor MAS items.

Learning materials.—In order to manipulate the variable of task difficulty two lists of verbal stimuli differing in meaningfulness as defined by Noble (**7**) were used. Table 1 presents the two 17-item lists employed. The mean *m* value for the high-meaningful list was 6.80, and for the low-meaningful list, 1.34.

Procedure.—Several weeks prior to participating in the verbal learning, all *Ss* had taken the two anxiety scales. The *Ss* were given no basis for assuming a relationship between the anxiety scales and the verbal learning.

Each list of 17 dissyllable words was exposed on a memory drum with a 2-sec. exposure per word and 6-sec. between trials. Each *S* received 25 trials on the appropriate list.[3] Learning was by the anticipation method, asterisks served as cues for anticipating the first word, and *Ss* were never informed about the criterion being used to determine the termination of the learning task. The *E* was unaware of *Ss*' anxiety scores at the time of running the verbal learning study.

The following instructions were given to all *Ss* upon entering the experimental situation: "You will see words appearing in this opening one at a time. After a word is presented, call out the next one before it appears. Of course, the first time through the list you won't be able to anticipate any words, but after that call out the word before it appears in the window. Prior to the beginning of each new trial there will be a short rest in which you will see blank spaces in the opening. Asterisks will indicate that the first word in the list will appear next. When you see them call out the first word. Do you understand?"

The Standard instructions groups received only those instructions; the Reassurance instructions groups received in addition the following statement from *E* prior to being administered the standard instructions: "Before we start, perhaps I could mention a few things that will be helpful to you in the learning of the list I am going to show you on this memory drum. Many people get unduly upset and tense because they do not learn the list in just a few trials. If you don't worry about how you are doing but rather just concentrate on the list, you will find you learn much more easily. These kinds of lists are hard and so it's no surprise or matter of concern if you progress slowly at first and make mistakes."

A person other than *E* was responsible for assigning *Ss* to the experimental treatment.

RESULTS

The data for the 25 trials were grouped into 5 blocks of 5 trials each. The score used in the analyses of variance was the number of correct responses per block.

[3] In those cases where *S* had five consecutive perfect trials *E* ended the experimental session at the end of the fifth consecutive perfect trial and *S* was assigned perfect scores for the remainder of the 25 trials.

TABLE 2

MEAN NUMBER OF CORRECT ANTICIPATIONS FOR 5 BLOCKS OF 5 TRIALS EACH FOR *Ss* CATEGORIZED ON BASIS OF TEST ANXIETY SCALE (TAS)

Trial Block	High TAS				Low TAS			
	Reassurance		Standard		Reassurance		Standard	
	High Meaning	Low Meaning	High Meaning	Low Meaning	High Meaning	Low Meaning	High Meaning	Low Meaning
1	17.38	9.88	20.88	5.75	16.75	3.88	23.63	5.00
2	40.50	25.13	44.63	13.00	34.50	14.88	52.88	19.13
3	60.50	36.38	66.63	23.38	56.38	27.50	72.00	36.38
4	75.25	49.25	78.13	33.38	64.00	36.00	81.63	51.25
5	82.13	54.63	79.38	38.25	71.38	43.88	80.75	63.88

Test Anxiety Scale.—Table 2 presents the mean number of correct anticipations for the eight experimental groups for the five blocks of trials. Table 3 presents the results of the $2 \times 2 \times 2 \times 5$ repeated measurement analysis of variance performed on the TAS data. As might be expected, the F for meaningfulness was highly significant ($P < .001$) with a greater number of correct anticipations given to the high- than to the low-meaningful lists. As expected, a significant F ($P < .01$) was obtained for the TAS × Instructions interaction. As Table 1 indicates, high-TAS *Ss* performed at a higher level under the Reassurance than under the Standard instructional condition. Low-TAS *Ss*, on the other hand, performed at a higher level under the Standard than under the Reassurance conditions. Thus, the differential instructions had opposite effects for high- and low-TAS *Ss*.

TABLE 3

SUMMARY OF TAS ANALYSIS OF VARIANCE OF CORRECT ANTICIPATIONS FOR 5 BLOCKS OF 5 TRIALS EACH

Source	*df*	*MS*	*F*
Test Anxiety (TAS)	1	.31	—
Meaningfulness (M)	1	55,862.45	77.25***
Instructions (I)	1	973.01	1.35
TAS × M	1	117.61	—
TAS × I	1	5,445.00	7.53**
M × I	1	1,757.81	2.43
TAS × M × I	1	649.75	—
Error (b)	56	723.01	
Trial Blocks (T)	4	28,422.19	460.88***
TAS × T	4	23.37	—
M × T	4	993.87	16.12***
I × T	4	26.50	—
TAS × M × T	4	162.99	2.64*
TAS × I × T	4	2.11	—
I × M × T	4	117.13	1.90
TAS × M × I × T	4	366.83	5.95***
Error (w)	224	61.67	

* $P = .05$.
** $P = .01$.
*** $P = .001$.

The Trials × Anxiety × Meaningfulness interaction ($P < .05$) was largely attributable to differences between high- and low-TAS groups on the high- and low-meaningful material for the fifth block of trials. On that block for high-meaningful groups, high-TAS *Ss* were superior to low-TAS *Ss*. For low-meaningful material on the same block, low-TAS *Ss* were superior to high-TAS *Ss*.

The significant Trials × Meaningfulness interaction ($P < .001$) indicates the learning curve for the high-meaningful material to be more steep than the curve for low-meaningful material early in learning and less steep late in learning.

Manifest Anxiety Scale.—The results of the $2 \times 2 \times 2 \times 5$ analysis of variance performed using MAS as the measure of anxiety revealed only two significant F's—those for Meaningfulness ($P < .001$) and for Trials ($P < .001$). Again performance was on a higher level under the high-meaningful list than under the low-meaningful list.

Discussion

In general, the results obtained in this experiment tended to support the stated expectations. There was no tendency toward an interaction between anxiety and instructions using MAS as the measure, but this interaction was significant for TAS.[3]

This finding would appear compatible with the assumption that high-anxious *Ss* bring with them into situations certain response tendencies which may interfere with performance. The reassurance condition seemed to reduce the effect of these interfering responses.

How might one account for the detrimental effect of reassurance for low-TAS *Ss* for both levels of meaningfulness? One fairly obvious possibility is that the reassurance for low-TAS *Ss* functioned to reduce their ego involvement in the task on which they were performing. If *S* is told that errors will inevitably occur, one conclusion which some *Ss* might have drawn is that efforts at reducing errors will be of little value. Thus, for low-TAS *Ss*, reassurance may have been a cue to "take it easy."

The present findings suggest that the more specific the psychometric measures of anxiety is to the class of situations in which we may be interested the more useful apparently will the anxiety measure be.

An interaction between anxiety and task difficulty of the type reported by Montague (6) was not obtained in this study. However, the Trials × Anxiety × Meaningfulness interaction and the four-factor interaction obtained for the TAS analysis suggest that difficulty of task and anxiety may be complexly related to each other.

[3] It would be of considerable interest to replicate the present study using very extreme high- and low-anxiety groups. On the basis of the present results and our interpretation of anxiety, we would expect that using high- and low-TAS groups more extreme than those used in this experiment would increase this interaction effect.

Summary

Effects on performance in verbal learning were measured for (*a*) anxiety as defined by Test Anxiety Scale scores and Taylor MAS scores (Bendig revision), (*b*) meaningfulness of material (Noble's *m* values), and (*c*) experimental instructions. The experimental instructions were: (*a*) standard instructions (*S* merely informed of the required manner of response to the memory drum), and (*b*) reassurance instructions (*S* informed that many errors were to be expected on the type of task on which he was performing and that he should not worry about making many errors). In a repeated measurement analysis of variance design the effects of the three experimental variables and their interactions with trials were assessed.

The results showed that for Test Anxiety, there was a significant interaction between Test Anxiety (high and low) and the kind of instructions (reassurance and standard) administered to *S*. It was found that the reassurance instructions were relatively facilitative for high Test Anxiety *Ss* and relatively detrimental for low Test Anxiety *Ss*. Under the standard instructional condition the performance of low Test Anxiety *Ss* appeared to be facilitated whereas just the opposite seemed to be true for high Test Anxiety *Ss*. When the data were analyzed in terms of Taylor MAS, there were no effects involving anxiety which attained significance. The results were discussed in terms of an interfering response interpretation of anxiety.

References

1. Bendig, A. W. The development of a short form of the Manifest Anxiety Scale. *J. consult. Psychol.*, 1956, **20**, 384. (Brief Report)
2. Child, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–171.
3. Farber, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.

4. Lucas, J. D. The interactive effects of anxiety, failure, and intra-serial duplication. *Amer. J. Psychol.*, 1952, **65**, 59–66.
5. Mandler, G., & Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
6. Montague, E. K. The role of anxiety in serial rote learning. *J. exp Psychol.*, 1953, **45**, 91–96.
7. Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.
8. Raphelson, A. C. The relationships among imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *J. abnorm. soc. Psychol.*, 1957, **54**, 13–18.
9. Saltz, F., & Hoehn, A. J. A test of the Taylor-Spence theory of anxiety. *J. abnorm. soc. Psychol.*, 1957, **54**, 114–117.
10. Sarason, I. G. Effect of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253–260.
11. Sarason, I. G. Effect of anxiety and two kinds of motivating instructions on verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54**, 166–171.
12. Sarason, I. G. The effect of anxiety and two kinds of failure on serial learning. *J. Pers.*, 1957, **25**, 383–392.
13. Sarason, I. G. Test anxiety, general anxiety, and intellectual performance. *J. consult. Psychol.*, 1957, **21**, 485–490.
14. Sarason, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. abnorm. soc. Psychol.*, 1958, **56**, 339–344.
15. Sarason, I. G. Intellectual and personality correlates of test anxiety. *J. abnorm. soc. Psychol.*, in press.
16. Waterhouse, I. K., & Child, I. L. Frustration and the quality of performance. III. An experimental study. *J. Pers.*, 1953, **21**, 298–311.

(Received January 17, 1958)

Journal of Experimental Psychology
Vol. 56, No. 6, 1958

THE RELATIONSHIP OF STIMULUS SIMILARITY AND NUMBER OF RESPONSES[1]

JACK RICHARDSON

Harpur College

The experimental tasks reported as examples of concept formation vary in many respects but seem to have two common characteristics which also serve to distinguish them from rote-learning tasks. The same response is correct for more than one stimulus and the stimuli which are designated by identical responses are discriminatively different but are "similar" in some way (have common elements, belong to the same class, etc.) while the stimuli designated by different responses are less similar.

Although verbalization of the basis of classification is often used as the criterion of concept attainment, there are two major methods of presenting concept examples and each has another criterion of concept learning or attainment. The first method presents single examples of each concept until the correct responses are learned and then presents a different example of each concept and requires the same response (e.g., **1, 7, 9**). The criterion of concept learning is the tendency for new examples of a concept to evoke the same response as the previous examples. This type of concept study resembles rote-learning studies of transfer in paired associates where the stimuli are similar and the responses identical (e.g., **4, 15**). Positive transfer from A-B to A′-B is expected in both, but A and A′ are examples of the same concept in the concept studies while in the transfer studies A and A′ have been rated as similar.

The second method presents several examples of a concept without requiring the correct response for the previous examples (e.g., **3, 10, 11**). The criterion for concept learning is the tendency for each example of the concept to evoke the same response. In these studies it would be possible to learn the correct response to each concept example independently, but it is assumed that assigning the same response to each example of a concept facilitates learning. Apparently there is positive transfer between examples of the same concept. To apply this method to rote-learning materials would require lists in which similar stimuli are paired with the same response.

There is some indication that similarity ratings are useful in predicting the relative difficulty of concepts (**13**). If stimulus similarity and number of responses are relevant to the differences between rote learning and concept learning then it should be possible to produce results characteristic of concept learning by manipulating rote-learning materials. The results should conform to the laws of transfer in verbal learning and at the same time show the faster learning and apparently better retention of concepts (**10, 11, 12**).

The present study is concerned with paired-associate lists in which the

[1] This research was supported by Grant G2594 from the National Science Foundation to the Research Foundation of State University of New York. Completion of this study was made possible by the cooperation of Paul Gimmie, Charles Traphagen, Malcolm Piester, and the students of Vestal Central School who served as *Ss*.

number of stimuli were held constant while the similarity of the stimuli and the number of different responses were varied. Of course, the important thing is the relationship between the stimulus similarity and the number of responses rather than either isolated from the other. One of the lists resembles the second method of concept presentation in that highly similar stimuli are designated by the same response while there are different responses for low-similarity stimuli.

METHOD

Lists.—Nine lists of 16 pairs were constructed with adjectives from Haagen (**6**) as stimuli and nonsense syllables from Glaze (**5**) as responses. The only differences in the lists were the number of *different* syllables used as responses and the similarity of the stimuli which were paired with a response. The design may be considered factorial with three different numbers of responses and three stimulus similarity relationships. Low-similarity adjectives were used as stimuli for three of the lists and consisted of 16 adjectives with no apparent or rated similarity. The high-similarity adjectives consisted of four groups of four highly similar items and were used as stimuli in six of the lists. The nonsense syllables were of either 47% or 53% association value, of minimum formal similarity, and either 2, 4, or 8 different syllables were correct as responses for all 16 stimuli in a list.

The nine lists are presented in Table 1 with the high-similarity stimuli in Column 1, the low-similarity stimuli in Column 2, and the responses which were paired with the stimuli in Columns 3–8. The headings of the columns containing the responses designate the different lists with the number indicating the number of different responses in the list and the letters indicating that the responses in that column were paired with the low (L) or the high-similarity (Ha and Hr) stimuli. For each different number of responses there were two pairings with the high-similarity stimuli. In one there was high intraresponse stimulus similarity (Ha), i.e., the responses were assigned so there was maximum similarity between stimuli paired with the same response. In the other there was high interresponse stimulus similarity (Hr), i.e., there was maximum similarity between stimuli paired with different responses.

It should be noted that only in the lists with four responses is it possible to assign a different response to each group of similar adjectives or to assign a single adjective from each similarity group to one response.

Procedure.—The 16 pairs in each list were arranged in three random orders and presented with a 16-mm. strip-film projector. The pairs were flashed from the back of a screen so that *S* could not see either *E* or the projector during

TABLE 1

LISTS

Item	Stimuli		Responses					
	High	Low	2-L 2-Ha	2-Hr	4-L 4-Ha	4-Hr	8-L 8-Ha	8-Hr
1	DOUBLE	VERBAL	BIH	BIH	BIH	BIH	BIH	BIH
2	DUAL	TWOFOLD	BIH	ZET	BIH	VOM	BIH	SUJ
3	PAIRED	HALLOWED	BIH	BIH	BIH	RUK	SUJ	VOM
4	TWOFOLD	DARING	BIH	ZET	BIH	ZET	SUJ	NAC
5	SACRED	PERFECT	BIH	BIH	VOM	BIH	VOM	BIH
6	DIVINE	DIRTY	BIH	ZET	VOM	VOM	VOM	SUJ
7	HALLOWED	REMOVED	BIH	BIH	VOM	RUK	NAC	VOM
8	HOLY	RISING	BIH	ZET	VOM	ZET	NAC	NAC
9	SPOKEN	ROBUST	ZET	BIH	RUK	BIH	RUK	RUK
10	ORAL	FLUID	ZET	ZET	RUK	VOM	RUK	YAD
11	VERBAL	PLAYFUL	ZET	BIH	RUK	RUK	YAD	ZET
12	VOCAL	GLEAMING	ZET	ZET	RUK	ZET	YAD	QIX
13	TRANQUIL	URGENT	ZET	BIH	ZET	BIH	ZET	RUK
14	PEACEFUL	OPEN	ZET	ZET	ZET	VOM	ZET	YAD
15	QUIET	STEADY	ZET	BIH	ZET	RUK	QIX	ZET
16	SERENE	LITTLE	ZET	ZET	ZET	ZET	QIX	QIX

the presentation. The stimulus of each pair was presented for 2 sec. followed by a 2-sec. presentation of the stimulus and response together. There was a 4-sec. intertrial interval.

Twenty-four different *Ss* learned each list; 12 *Ss* were college students from introductory psychology courses and 12 were junior or senior high school students. The *Ss* from each of the two groups of students were assigned to lists as they appeared at the laboratory. None of the *Ss* had previously served in a learning experiment. They were required to spell the syllables and were given the usual instructions for paired-associate learning by the anticipation method. They were not told that the same response would be correct for more than one stimulus or that the stimuli were grouped in any way for the assignment of responses.

To avoid selection due to loss of *Ss* and to approximately equate performance at the end of the first session, the lists were presented for a specific number of trials. The lists with two different responses were presented for 5 trials, those with four different responses for 10 trials, and the lists with 8 responses for 15 trials. After learning for the appropriate number of trials, *Ss* were dismissed and returned 48 hr. later for the second session. Since differential transfer was expected between items in the lists, it was important that the responses should not be presented during any test of recall. To control for intralist transfer and for warm-up, the first five trials of the second session were recall trials. During the five recall trials the lists were presented in exactly the same way as during learning except that the responses were not presented and there was no indication to *S* whether a response was right or wrong. The *Ss* were told that the responses would not appear for a few trials but that they should continue to spell as many responses as possible each time through the list. After the five recall trials the responses were again presented in the usual fashion and relearning continued to a total of 25 trials for both sessions or to a criterion of three successive perfect trials. For purposes of the analysis it was assumed that all responses would be correct after three successive perfect trials.

Results

Learning.—The mean number of correct responses on Trial 25 of each list is given in Table 2 and the analysis of variance is presented in Table 3. The college students are superior to the high school students, but there is no significant interaction of students with the other variables. The two groups are combined in Fig. 1 and the mean number of correct responses on each trial is presented for each list. The first break in each curve indicates the 48-hr. retention interval, and the second the end of the five recall trials when the responses were again presented.

Number of different responses is highly significant, but this analysis does not take into account the differential probability of guessing the correct response. The appreciable number of times *Ss* did not respond makes the value of a correction for guessing doubtful. However, the correction used by Bricker (**2**) was ap-

TABLE 2

Mean Number of Correct Responses on Trial 25

Similarity	College Students			High School Students		
	No. of Responses			No. of Responses		
	2	4	8	2	4	8
Hr	12.08	9.17	8.17	11.25	8.50	8.92
L	14.17	12.00	10.92	11.25	11.17	11.50
Ha	14.17	15.33	12.00	13.67	13.17	7.83
Combined	13.47	12.17	10.36	12.06	10.94	9.42

TABLE 3

Analysis of Variance for the Number of Correct Responses on Trial 25

Source	*df*	*MS*	*F*
Students (St)	1	77.04	7.48*
No. of Responses (NR)	2	150.04	14.58*
Similarity (Sm)	2	173.51	16.86*
NR × Sm	4	37.83	3.67*
St × Sm	2	18.76	1.82
St × NR	2	1.01	.10
St × NR × Sm	4	21.17	2.06
Within	198	10.29	

* $P = .01$.

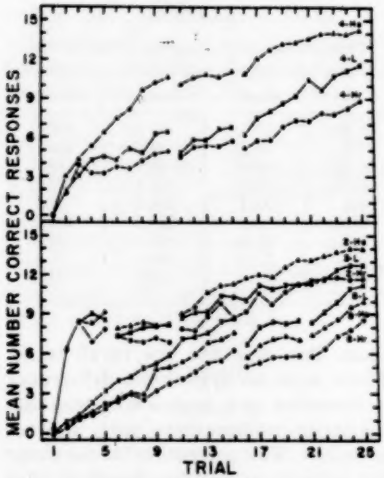


FIG. 1. Mean number of correct responses on each list as a function of trials.

plied to Trial 25 and an analysis of variance was computed on the corrected scores. This decreased the *F* for number of responses to .97 and reduced the *F* for the interaction between number of responses and similarity to 2.98. The *F*'s for Similarity and for Students increased to 17.51 and 8.82, respectively. This correction probably overestimates the effect of guessing, but there are other indications that number of responses may not be an effective variable in learning. In Fig. 1 it is apparent that the performance level expected from guessing alone is attained in approximately the same number of trials for lists with different numbers of responses and that the differences between the 2-, 4-, and 8-response lists tend to decrease after this point. There is also a significant interaction between number of responses and similarity with the smallest differences between the low-similarity lists. In view of this, the significant effect of number of responses is attributed to guessing and to similarity relationships.

The over-all effect of similarity is highly significant with the high interresponse stimulus similarity lists being more difficult, and the high intraresponse stimulus similarity lists less difficult, than the low-similarity lists. However, the significant interaction indicates that the effect of similarity is not independent of the number of responses. Similarity as it is used here is a dimension only if considered in relation to the responses. If the number of stimuli remains constant then the relationships in the Ha and Hr lists must change as the number of responses change. If we consider intralist transfer as a function of the similarity of the stimuli then there would be little or no transfer in the low-similarity lists while each stimulus in the high-similarity list would produce transfer to three others. The direction of transfer would depend on whether the three stimuli had the same or different responses as the stimulus which is producing the transfer. Thus the

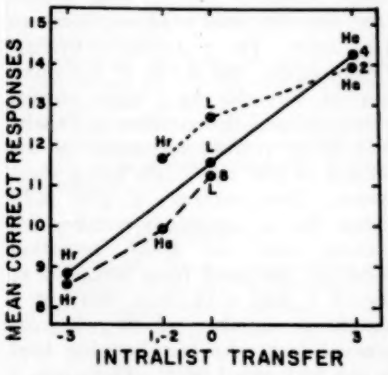


FIG. 2. Mean number of correct responses on Trial 25 as a function of the number of items within a similarity group to which the transfer would be positive or negative.

high-similarity lists would have either negative transfer to three items within a similarity group (Lists 4-Hr and 8-Hr), positive transfer to one and negative to the other two (Lists 2-Hr and 8-Ha), or positive transfer to three items (Lists 2-Ha and 4-Ha). In Fig. 2 the number of correct responses on Trial 25 is plotted as a function of the amount of transfer expected. The abscissa indicates the number of items, within a single similarity group, to which transfer is expected to be positive or negative. There is a separate curve for each of the three different number of responses and the similarity is indicated at each point. It is clear that the number of correct responses increases as the number of items to which positive transfer is expected increases. This effect is not dependent on the number of different responses in the list.

Retention.—In Fig. 1 it is apparent that there was an appreciable loss on all lists from the last learning trial to the first recall trial and, even though *S*s were given no information as to the correctness of their responses, there was a gain over the five recall trials on all except List 2-Hr. For a statistical evaluation a count was made of the number of *S*s who had more correct responses and the number of *S*s who had fewer correct responses on the second of the two trials being compared. Computation of χ^2's indicated that a significant number decreased over the 48-hr. retention interval, increased from Recall 1 to Recall 5, and that there was not a significant number of *S*s who decreased from the last learning trial to the fifth recall trial. There was a tendency for more loss over the retention interval in the low-similarity lists but there was also more gain in these lists over the five recall trials. There were no systematic differences in retention as a function of stimulus similarity relationships with the responses. The percentage of the number correct on the last learning trial which was correct on the first and fifth recall and the first relearning trial is given for each list in Table 4. It should be noted that retention is uniformly high and that it increases from Recall 1 to Recall 5.

TABLE 4

PERCENTAGE OF RETENTION AFTER 48 HR.

List	Recall 1	Recall 5	Relearn 1
2-Hr	93.1	87.2	100.5
2-L	86.3	95.6	102.0
2-Ha	86.3	89.0	98.6
4-Hr	91.5	117.8	105.1
4-L	76.1	106.4	98.1
4-Ha	94.9	104.3	102.4
8-Hr	80.5	93.3	92.0
8-L	77.8	98.6	96.6
8-Ha	75.7	92.8	98.9

Errors.—The errors on the last learning trial, the first and fifth recall trials, and the first and second relearning trials were classified as omissions (failure to make an overt response), errors from within the list, or others which included partial responses and combinations of letters which were not correct for any item in the list. There were no systematic

TABLE 5

TOTAL NUMBER OF ERRORS IN EACH CLASSIFICATION FOR ALL LISTS COMBINED

Type of Error	Last Learning Trial	Recall 1	Recall 5	Relearn 1	Relearn 2
Omissions	545	587	433	586	395
Within	1084	1183	1130	1104	1155
Other	146	244	244	96	93
Total	1775	2014	1807	1786	1643

differences in the distribution of types of errors as a function of similarity. Although there were many more errors in the Hr groups, the percentage of these that were from within the list was no greater than in the Ha groups. The number of different types of errors in the five trials for all lists combined are given in Table 5. The increase in correct responses over the five recall trials is primarily the result of the decrease in the number of omissions.

Discussion

It is apparent from Fig. 2 that rote-learning transfer principles predict the relative speed of learning when similarity relationships are varied in relation to the responses. The higher the similarity between stimuli with the same response the faster the learning, and the higher the similarity between stimuli with different responses the slower the learning. The positive intralist transfer commonly found in concept learning can be produced by manipulating stimulus similarity. List 4-Ha is most comparable to the concept lists since there is one response for each group of similar stimuli. It is here that the difference between the Ha list and the L list is greatest.

Retention does not seem to be related to similarity and this agrees with other studies in which stimulus similarity has been varied (**14**). However, the retention is high. If one considers only the 4-Ha list, there is 94.9% retention on the first recall trial and this increases to 104.3% on the fifth recall trial. Differences in number of items, degree of learning, retention interval, and procedure prohibit direct comparison with concept studies (**10, 11, 12**), but the retention of concepts is not obviously better. In view of the high degree of retention in this study it seems that a direct comparison of concept and rote materials which used the same procedure and equated the degree of learning might show no differences in retention.

The five recall trials during which the responses were not presented resulted in an increase in the number of correct responses on all lists except 2-Hr and the small increase in the 2-response lists may be due to the fact that the performance at the end of learning was at approximately the guessing level. This procedure seems similar to warm-up before recall found by Irion (**8**) but differs in that it is practice on the same task rather than some other task. On the other hand, it cannot be considered learning in the usual sense because there was no information given as to the correctness of the response. In Table 4 there seems to be little relation of the percentage of increase over the five recall trials with either similarity or number of responses. The increase is definitely related to the number of trials on recall, but the limit cannot be determined from this experiment. Further experimentation is required to relate this phenomenon to warm-up and to determine what aspects of the procedure are necessary to produce it.

Summary

This study varied the number of different responses and the similarity of the stimuli in relation to the responses in paired-associate lists. In 16-item lists either 2, 4, or 8 nonsense syllables were used as responses and either high- or low-similarity adjectives were used as stimuli. The similarity of the stimuli which had the same response was varied.

It was found that similarity was a variable in learning and that the list with a single response for each group of similar adjectives showed high positive intralist transfer characteristics of concept learning. Retention did not seem to be related to similarity, but was uniformly high.

References

1. Baum, M. H. Simple concept learning as a function of intralist generalization. *J. exp. Psychol.*, 1954, **47**, 89–94.
2. Bricker, P. D. The identification of redundant stimulus patterns. *J. exp. Psychol.*, 1955, **49**, 73–81.
3. Buss, A. H. A study of concept formation as a function of reinforcement and stimulus generalization. *J. exp. Psychol.*, 1950, **40**, 494–503.

4. Gibson, E. J. Retroactive inhibition as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28,** 93–115.
5. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35,** 255–267.
6. Haagen, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27,** 453–463.
7. Heidbreder, E. The attainment of concepts: VIII. The conceptualization of verbally indicated instances. *J. Psychol.*, 1949, **27,** 263–309.
8. Irion, A. L. Retention and warming-up effects in paired-associate learning. *J. exp. Psychol.*, 1949, **39,** 669–675.
9. Hull, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28,** No. 1 (Whole No. 123).
10. Oseas, L., & Underwood, B. J. Studies of distributed practice: V. Learning and retention of concepts. *J. exp. Psychol.*, 1952, **43,** 143–148.
11. Reed, H. B. Factors influencing the learning and retention of concepts: I. The influence of set. *J. exp. Psychol.*, 1946, **36,** 71–87.
12. Richardson, J. Retention of concepts as a function of the degree of original and interpolated learning. *J. exp. Psychol.*, 1956, **51,** 358–364.
13. Richardson, J., & Bergum, B. O. Distributed practice and rote learning in concept formation. *J. exp. Psychol.*, 1954, **47,** 442–446.
14. Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45,** 143–149.
15. Yum, K. S. An experimental test of the law of assimilation. *J. exp. Psychol.*, 1931, **14,** 68–82.

(Received January 29, 1958)

Journal of Experimental Psychology
Vol. 56, No. 6, 1958

FREE RECALL OF REDUNDANT STRINGS OF LETTERS [1]

GEORGE A. MILLER

Harvard University

The well known fact that meaningful verbal material is easier to memorize than nonsense material is usually explained by saying that meaningful material permits more positive transfer from previous learning; *S* has already learned a great deal about the meaningful material before the experiment begins. Exactly *what* the *S* has already learned that he can transfer to the memorization of meaningful material is usually not specified, however. The positive transfer may be attributable to something about meaning per se that makes the task easier, or it may be attributable to the fact that meaningful verbal material usually obeys certain familiar rules of formation, or both. In order to study this question, Miller and Selfridge (**5**) used statistical approximations to English which *S*s were required to reproduce after one hearing; they came to the conclusion that familiarity with the rules which govern the sequential properties of meaningful passages is probably much more important in facilitating recall than is *S*'s understanding of the passage. This conclusion implies that we should take a closer look at the effects of nonrandom rules of formation.

Aborn and Rubenstein (**1, 6**) initiated such an exploration of the effects of rules in a series of studies that used sequences of nonsense syllables. The sequences were constructed according to rules that varied from completely random to highly deterministic. In the terminology of information theory (**7**), the sequences formed according to elaborate and restrictive rules were more redundant, carried less information per syllable, than the random sequences. It was found that, for *S*s familiar with the rules of formation, the redundant sequences were easier to memorize, but not as much easier as would be predicted on the hypothesis that the amount of information acquired is invariant under changes in the rules of formation. These results confirm and extend the conclusion drawn by Miller and Selfridge and do so in a situation where meaningful words are eliminated and *E* has complete knowledge of the rules of formation and of *S*'s prior experience with those rules.

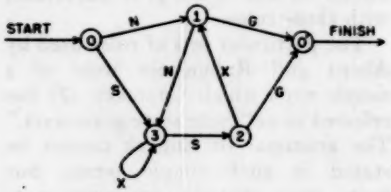
The particular sets of rules used by Aborn and Rubenstein were of a simple type which Chomsky (**2**) has referred to as "finite state grammars." The grammar of English cannot be stated in such simple terms, but finite state systems are quite convenient for the experimental study of the effects of rules on memorization. Chomsky and Miller (**3**) have described the formal properties of finite state languages and their results are used below to describe the materials in the present study. In spite of their formal simplicity, these finite state systems pose some complicated psychological problems. We would like to know what changes, if any, in a set of rules leave performance

[1] This research was supported under Contract AF 33(038)-14343 and appears as report number AFCRC-TN-58-50 (ASTIA Document Number AD 146780). Reproduction for any purpose of the U. S. Government is permitted. The author is indebted to Elizabeth A. Friedman for the collection of the data.

invariant, and so we are faced with a large and tangled set of possible relations between the formal aspects of the finite state systems and the empirical aspects of *S*'s behavior. The present study does little more than illustrate the complexity of this formidable task.

Characterization of Redundant Strings

All strings of symbols used in these experiments were composed of four letters of the alphabet (G, N, S, and X) chosen randomly from the consonants. With these four letters it is possible to form 4^{λ} different strings of letters of Length λ. A finite state generator is an algebraic system that selects a particular subset, the well-formed strings, out of the set of all possible strings (**2, 3**). The particular finite state generator used here is displayed in graphical form in Fig. 1, where the numbered circles represent the states of the system, the arrows represent the permissible transitions from state to state, and the letters associated with each arrow represent the symbols generated in moving from one state to the next. All well-formed strings are generated by starting in State 0 and following any path through the diagram that ends on the first visit to State 0′. Thus, the generator selects the well-formed strings: NG, SSG, NNSG, SSXG, SXSG, NNSXG, SXXSG, etc. At the bottom of Fig. 1 the rules of transition are stated explicitly; for example, (0, N, 1) means that when the system is in State 0 it can generate the symbol N and thus move next to State 1. A state is defined as the set of all rules applicable in that state; for example, State 2 consists of two rules, (2, G, 0′) and (2, X, 1).



RULES OF TRANSITION:

(0,N,1)	(1,N,3)	(2,G,0′)	(3,S,2)
(0,S,3)	(1,G,0′)	(2,X,1)	(3,X,3)

Fig. 1. Diagram of the finite state generator of the redundant strings used in this experiment. An admissible string is any sequence of letters generated by starting in State 0 and ending on the first arrival at State 0′. An equivalent, algebraic description of the generator is given in the rules of transition beneath the diagram; each rule gives the state of the system in which the rule is applicable, the symbol generated in transition, and the state in which the system is left after the rule is applied.

For this particular generator, the number $F(\lambda)$ of strings of Length λ or less can be computed recursively from the equation $F(\lambda) = 2F(\lambda - 1) - F(\lambda - 2) + F(\lambda - 3) - F(\lambda - 4)$, where $F(1) = 0$, $F(2) = 1$, $F(3) = 2$, and $F(4) = 5$.[3] The largest real root of this finite difference equation is 1.4656, and for large λ, $F(\lambda)$ approaches $1.731(1.4656)^{\lambda}$. If we use Shannon's (**7**) measure of channel capacity, this system has a capacity of

$$\lim_{\lambda \to \infty} \frac{\log_2 F(\lambda)}{\lambda} = \log_2 1.4656 = 0.552 \text{ bit/letter.}$$

By comparison, the set of all strings formed on this set of four letters has a capacity of $\log_2 4 = 2$ bits/letter. Thus, the redundancy is $1 - (0.552/2) = 0.724$, or 72.4%. This can be interpreted to mean that the well-formed strings are 3.62 times as long as they would need to be in order to encode the same amount of information.

It is possible to show that such a finite state generator can be decomposed, or "factored," into a set of simpler generators. In the present example, we first find all those paths between States 0 and 0′ which visit no state more than once. These paths generate the basic strings NG, NNSG, SSG, and SSXG. All well-formed strings can be formed by inserting loops at appropriate points in these four basic strings. We symbolize a loop by placing it in parentheses, where the second parenthesis indicates that we have returned to the same state we were in at the first parenthesis. Using this notation, we can summarize the entire set of well-formed strings: N(N(X)SX)G, N(N(X)SX)NSG, S((X)(SXN))SG, and S((X)(SXN))SXG. This method of representing the well-formed strings is more compatible with the way *S*s tend to organize them than are the representations in terms of graphs or transition rules. It should be noted that in this particular generator, the

[3] The number $f(\lambda)$ of strings of exactly length λ can be similarly computed from the equation $f(\lambda + 3) = f(\lambda + 2) + f(\lambda)$, for $\lambda = 2, 3, \ldots$, given $f(1) = 0$, $f(2) = f(3) = 1$, and $f(4) = 3$. Obviously, $f(\lambda + 1) = F(\lambda + 1) - F(\lambda)$. The amount of information per letter computed for $F(\lambda)$ is the same as that computed for $f(\lambda)$. (Cf. **3**, p. 111.)

TABLE 1

LISTS OF STRINGS USED

Redundant Strings		Random Strings	
L_1	L_2	R_1	R_2
SSXG	NNSG	GNSX	NXGS
NNXSG	NNSXG	NSGXN	GNXSG
SXSXG	SXXSG	XGSSN	SXNGG
SSXNSG	NNXSXG	SXNNGN	GGSNXG
SXXXSG	NNXXSG	XGSXXS	NSGNGX
NNSXNSG	NNXXSXG	GSXXGNS	NGSXXNS
SXSXNSG	NNXXXSG	NSXXGSG	NGXXGGN
SXXXSXG	SSXNSXG	SGXGGNN	SXGXGNS
SXXXXSG	SSXNXSG	XXGNSGG	XGSNGXG

basic strings themselves seem to group together; NG is included in NNSG and SSG is included in SSXG, and the *S*s seemed to distinguish only two basic types of strings. The *S* who notices these similarities among the strings can use them to reduce the amount of information he is required to remember.

METHOD

The generator in Fig. 1 gives 22 well-formed strings of Length 7 or less. Of these 22, 18 were chosen in such a way that two lists of nine strings, each containing the same number of strings of the same length, could be constructed. The assignment of a string to one list or the other was determined by a table of random numbers. The result of this procedure was the two lists, L_1 and L_2, in the first two columns of Table 1.

In addition, Table 1 contains two further lists, R_1 and R_2, of nine strings composed of the same four letters. Lists R_1 and R_2, however, were constructed with the aid of a table of random numbers. The only constraint in their formation was the requirement that the R lists should have the same number of strings of each length as did the L lists.

These strings were printed in block letters on 3 × 5-in. cards. The *S* was seated across the table from *E*, who laid the cards face up in front of *S* at a rate of one card every 5 sec. After all nine cards in the list had been presented, *E* picked up the deck of cards and shuffled it while *S* wrote, in any order, all the strings he could recall. When *S* finished, *E* removed the response sheet and proceeded to the next trial. Ten such trials were given and the order of the cards was varied haphazardly on each trial. Each *S* studied two lists, and at the end of the tenth trial on the second list he wrote all the strings he could recall from the first list.

The *S*s were undergraduate and graduate students at Harvard and Radcliffe Colleges and members of the staff of the Harvard Psychological Laboratories. The original design called for two *S*s each to learn the lists L_1L_2, L_2L_1, R_1R_2, and R_2R_1, in that order, and one *S* to learn L_1R_1, L_1R_2, L_2R_1, L_2R_2, R_1L_1, R_1L_2, R_2L_1, and R_2L_2, which would have given four *S*s on each type of pair: LL, RR, LR, and RL. The results with the RL condition were so variable, however, that it was run three times, giving 12 *S*s in all for this condition.

The *S*s were told nothing about the method of constructing the lists. The same explanation of the procedure to be followed was given to every *S*: he was to read the nine cards and write down as many strings as he could remember on each trial.

The response sheets were scored for the number of correct strings. Incorrect responses were ignored.

RESULTS

The mean number of correct responses is shown in Fig. 2 as a function of the trial number. The plot on the left gives the mean number correctly recalled by the eight *S*s who learned an L list first (open circles) and also by the 16 *S*s who

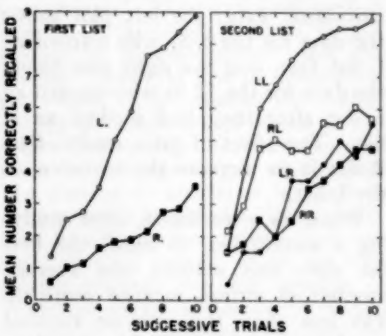


FIG. 2. Mean number correct as a function of the trial number. L indicates a redundant list and R indicates a random list. On the left, the curve labeled L. represents the results on the first list for the eight *S*s who studied lists LL or LR, in that order, and the curve labeled R. represents the 16 *S*s who received RL or RR. On the right, the curves labeled LL, LR, and RR are based on data for the second list studied by four *S*s each; the RL function is based on 12 *S*s.

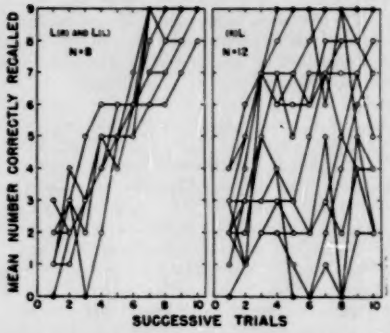


FIG. 3. Individual data for eight *Ss* learning a list of redundant strings initially are given in the graph on the left. On the right, the individual data are given for the 12 *Ss* who studied the redundant strings after they had previously studied a list of random strings.

learned an R list first (solid squares). The plot on the right in Fig. 2 gives the mean number correct on the second list for each of the four groups.

Averaging data from several *Ss* gives a poor picture of what happened in the RL condition. Figure 3 shows individual data: the left plot shows the data for the 8 *Ss* who learned an L list first, and the right plot shows the data for the 12 *Ss* who learned an L list after they had studied an R list. The effect of prior study of an R list is to increase the variance on the L list.

When *Ss* were asked, after studying a second list, to recall the first list they had studied, the average number of strings recalled correctly was less than the number recalled on the tenth trial with the first list. In the LL condition, the average recall dropped from 8.75 strings on the last trial with the first list to 5.0 strings on the retest. In the LR condition it dropped from 9.0 to 5.0. In the RL condition, it dropped from 3.33 to 1.83. And in the RR condition it dropped from 4.25 to 0.5.

DISCUSSION

It is clear that the redundant strings were learned more quickly than the random strings. One *S* in the LR group expressed the difference when he commented after the experimental session, "The first list was like studying, but the second was like cramming." The fact that *Ss* were organizing the strings into related clusters could be seen in the serial order of their recall on the later trials with the L lists.

It is often assumed, at least for nonsense materials, that associative interference will increase as the items to be learned are made more similar. This assumption may not be valid when the method of free recall is used, since Underwood (9) has shown that recall and intralist similarity are not systematically related. In the present experiment the redundant lists have greater intralist similarity than the random lists,[3] yet the similarities are an advantage rather than a handicap. The *Ss* avoided interference due to stimulus generalization by recoding the redundant strings. For example, List L_2 contains NNSG, NNSXG, NNXSXG, NNXXSG, NNXXSXG, and NNXXXSG. Given NNSG, these six strings are easily recoded as 00, 01, 11, 20, 21, and 30, where the two digits indicate the number of X's preceding and following the S. Then SSXNSXG and SSXNXSG can be learned together,

[3] The greater intralist similarity of the redundant strings was demonstrated by a search task. Lists of 30 and 60 redundant strings and 30 and 60 random strings were typed on separate sheets of white paper and 10 *Ss* were asked to search for particular strings. With 30 alternatives, the median search time was 15.5 sec. for redundant strings, but only 9.3 sec. for random strings. With 60 alternatives, the corresponding medians were 33.0 and 15.5 sec. It took about twice as long to search through the redundant strings, and it is in this sense that one can say the redundant strings have a greater degree of similarity. We are indebted to Sherry Blumenthal for collecting these data. It is interesting to note that Fitts, Weinstein, Rappaport, Anderson, and Leonard (4) also found that the recognition of redundant visual patterns was slower than the recognition of random patterns in a search task.

since one is derived from the other by a simple inversion of letters. That leaves SXXSG as the ninth string, the only one that doesn't "fit" with any other. By such grouping and recoding devices as this it is possible to reduce the difficulty of the task. Similarity of the strings is not a handicap if *S* does not choose to remember the strings in the form in which they are presented, but recodes them and remembers the coded form instead.

If we consider the quantities of information involved, we can ask whether the amount of information per trial is approximately the same for the redundant and the random strings. This would be a surprising result in any case, but especially in this experiment where *S*s had no prior knowledge of the rules underlying the redundant strings and could scarcely be expected to exploit them in an optimal fashion. The question is easily tested with the data, however. With random strings, the average *S* learned 3.63 strings on his first 10 trials and these strings contained an average of 20.5 letters. At 2 bits/letter, this corresponds to 41 bits of information, or 4.1 bits/trial for the random strings. With redundant strings, on the other hand, the average *S* learned all nine strings, or 54 letters, in his first 10 trials; at 0.552 bits/letter, this amounts to 29.8 bits, or approximately 3 bits/trial. These values, 4.1 and 3 bits/trial, actually underestimate the difference, since the computation for the random strings ignores the information *S* may have acquired about the strings he was unable to reproduce perfectly. This result confirms the conclusion of Aborn and Rubenstein (**1, 6**), who also found that with highly redundant strings *S*s learn more symbols but less information than with random strings.

In order to explore this question of information storage in more detail, eight *S*s learned a list of nine random strings composed of G, N, S, and X where the total amount of information in the new set of nine random strings was approximately the same as the total amount of information in the set of nine redundant strings. The total amount of information involved in the redundant strings was 29.8 bits, so the new random strings had to be composed of 15 letters at 2 bits/letter in order to give the equivalent 30 bits of information. These 15 letters were arranged in three strings of Length 1 and six strings of Length 2. The experimental procedure was the same as before. The average number correctly recalled on successive trials was 6.0, 7.6, 8.0, 8.0, 8.3, 8.5, 8.5, 8.6, 8.6, and 8.6 strings. If, in order to estimate the amount of information stored per trial, we assume that in three trials the average *S* had learned eight strings, or 26 bits, then we obtain an estimate of about 8.3 bits per trial. This is a crude estimate, but even so it is clear that the amount of information acquired per trial is not a constant independent of the way the information is encoded. The shorter random strings are more easily memorized and provide a more efficient method of storing information.

The total amount of information that *S* acquires in a series of trials with a given list can be broken up into at least four factors. The total amount of information acquired is equal to the number of letters learned times the average number of bits of information per letter. The number of letters learned is approximately equal to the number of strings correctly recalled times the average length of a string. The number of bits/letter is equal to the maximum possible bits/letter multiplied by a factor that depends upon the redundancy of strings. Thus the four factors involved are the number of letters per string, the number of strings, the size of the alphabet used, and the redundancy of the strings. Symbolically, the amount of information *S* has acquired on any given trial is not less than

$$(1 - R) \log_2 m \sum_{i=1}^{N} \lambda_i s_i$$

where m is the size of the alphabet used, R is the redundancy of the strings, λ_i

is the number of letters in the ith string, s_i is a function that assumes the value 1 if String i is recalled and 0 if String i is not recalled, and N is the number of strings. The independent variables under E's control are m, R, N, and λ_i, in addition to the number of the trial, and the dependent variable is s_i, the strings which S recalls.

With random strings, the evidence available indicates that m, the number of different letters, has relatively little effect on the number of trials required to reach a given criterion of mastery (**8**). R is, of course, zero for random strings, so the important variables are λ_i and N, the length of the strings and the length of the list of strings to be memorized. With redundant strings, on the other hand, the task may actually become easier as m increases, since the rules underlying the redundancy usually become more obvious. As R increases, learning will become easier for any S who takes advantage of the redundancy, but the conditions which enable him to use it in order to simplify his task remain to be explored. The effect of increasing N when the strings are redundant should be far less than when they are random, since the strings fall into certain recognizable classes defined by the basic strings and the number of basic strings does not vary as a function of N. The effect of λ_i is not obvious. It would be interesting to keep the total number of letters constant while N and λ_i varied inversely.

A principal question in such an analysis, however, is whether the measure of redundancy, R, is adequate to represent all the different kinds of redundancy that might lead to the same value of R. With respect to the finite state generators which determine the redundancy of the strings, the major variables to be explored include the number of different letters used, the number of states, the number of rules of transition, and the number of basic strings.

This quick glance at the parametric aspects of redundant strings makes their exploration look rather formidable, especially when we remember the large number of parameters we already have in studies of verbal learning. The prospect is further complicated by the necessity of exploring the conditions that facilitate the efficient discovery and use of the redundancies in such strings. This question raises problems more nearly akin to concept formation than to the usual studies of verbal learning. Although some psychologists seem to hold the opinion that concept formation is merely a complicated kind of verbal learning, the present analysis suggests the converse—namely, that verbal learning is merely a simple form of concept formation. In either view, however, it seems likely that the same problems will arise in both areas and that one will not progress without the other.

The design of the present experiment is appropriate for the study of transfer effects in verbal learning. In Fig. 2 there is apparent a positive transfer for the LL, LR, and RR conditions; the improvement may be simply another case of the well known "warm-up effect." The RL condition, however, shows negative transfer and Fig. 3 indicates that this negative transfer is due to a large increase in the variance between Ss. In the terminology of verbal learning experiments we should probably call this result a negative "transfer of principles," since the prior learning of a random list apparently caused some Ss to accept the "principle" that the new task was also random. It is just as natural, however, to speak of an *Einstellung* effect, or of functional fixedness. The strategy that Ss adopted for memorizing redundant strings apparently did not handicap them when they later encountered random strings, but the reverse was not true: The strategy adopted for the random strings was not efficient when later applied to redundant strings.

However, the *Einstellung* effect suggested by the increased variance in the scores obtained on redundant strings following prior experience with random strings may be peculiar to the particular finite state generator used in this experi-

ment. Subsequent studies with other generators have not shown this disruption so clearly. Until we have a better conception of how *S*s attempt to organize redundant strings, therefore, it is unwise to overgeneralize this feature in the results.

Summary

The present experiment explores the recall of redundant, as opposed to random, strings of symbols. First, an algebra for generating strings of symbols according to explicit rules is defined and a set of redundant strings is generated. A subset of these strings is memorized by *S*, using the method of free recall. As a control, the same procedure is used for equivalent strings generated by a table of random numbers. It is found that, although *S* knows nothing of the rules of formation, the redundant strings are more easily memorized. Since the redundant strings show greater intralist similarity, it is concluded that *S*s group and recode the redundant strings, thus avoiding the interference effects that would be expected for the uncoded strings.

In agreement with earlier studies, it is found that the amount of material learned increases when the strings are redundant, but the amount of information, measured in bits, decreases. In terms of the information measure, therefore, redundancy in the materials to be learned does not increase the efficiency of learning. The several parameters involved in these relations are briefly reviewed.

The results also indicate that, for the particular redundancy used, some *S*s find the redundant strings more difficult to learn if they have first learned a series of random strings.

References

1. Aborn, M., & Rubenstein, H. Information theory and immediate recall. *J. exp. Psychol.*, 1952, **44**, 260–266.
2. Chomsky, N. Three models for the description of language. *IRE Trans. Information Theory*, 1956, **IT-2**, 113–124.
3. Chomsky, N., & Miller, G. A. Finite state languages. *Information and Control*, 1958, **1**, 91–112.
4. Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, 1956, **51**, 1–11.
5. Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.
6. Rubenstein, H., & Aborn, M. Immediate recall as a function of degree of organization and length of study period. *J. exp. Psychol.*, 1954, **48**, 146–152.
7. Shannon, C. E. A mathematical theory of communication. *Bell System Tech. J.*, 1948, **27**, 379–423.
8. Simon, H. A. Amounts of fixation and discovery in maze learning behavior. *Psychometrika*, 1957, **22**, 261–268.
9. Underwood, B. J. Intralist similarity in verbal learning and retention. *Psychol. Rev.*, 1954, **61**, 160–166.

(Received February 3, 1958)

Journal of Experimental Psychology
Vol. 56, No. 6, 1958

TRAINING AND TRANSFER AS A FUNCTION OF COMPONENT INTERACTION[1]

GEORGE E. BRIGGS AND LAWRENCE K. WATERS

Ohio State University

In specifying the complexity level of a particular motor skill task, one finds that a major factor is the number of components or subtasks which define the whole task. Thus, as the number of components of a whole task is increased, *S* may find acquisition of the whole skill to be increasingly difficult. An obvious expediency here involves initial training on parts or single dimensions with subsequent practice on combinations of components as in the part-whole paradigm. Unfortunately, practice on individual components generally does not increase proficiency levels on the whole task to that level achieved by whole-task training itself (**1, 3, 5, 8**). However, these differences in performance between whole- and part-practice schedules at the time of transfer can be reduced rather quickly by continued post-transfer training on the whole task. This can result in less training required on the whole task itself following part-task practice, a saving which can be of considerable importance (**1, 8**).

Except for one study (**1**), these generalizations are based on experiments in which the components defining the whole task were physically independent and thus could be manipulated independently by *S*s. For example, when *S* makes a correction for an azimuth error with the Pedestal Sight Manipulation Apparatus, it is not necessary to coordinate simultaneous responses to the elevation or range dimensions of control, and the evidence suggests that *S* will indeed respond to each of the several components independently, on the average (**2, 7**).

Most skill systems involve components which are not physically independent, one from another; interactions exist among the components of the whole system such that corrections applied to one component affect other components as well and therefore necessitate concomitant responses by *S* to these other components. To illustrate, in the operation of an aircraft, a lateral deflection of the control column has the primary effect of introducing roll with a subsequent change in heading. However, this control action has the secondary effect of a reorientation of the lift vector with subsequent loss in altitude by the aircraft. Thus, the pitch and roll components of the system interact in such a way as to require *S* to pull back on his control column whenever he desired to displace that control laterally.[2] By such a coordinated response, *S* can thereby cancel the effects of the pitch-roll (P × R) interaction.

[1] This research was conducted in the Laboratory of Aviation Psychology of the Ohio State University and was supported in part by the United States Air Force under Contract No. AF 41(657)-70 with the OSU Research Foundation, monitored by Robert L. French and Bryce O. Hartman, who served as Scientific Officers for the AF Personnel and Training Research Center. Permission is granted for reproduction publication, use, and disposal in whole or in part by or for the United States Government.

[2] We assume no rudder corrections are required.

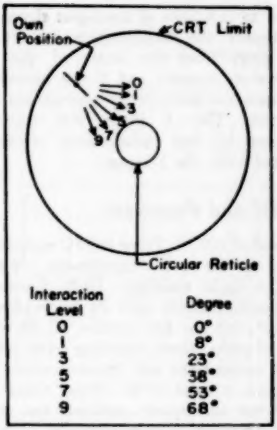


FIG. 1. The elements of the tracking display and the schematized effects of component interaction on the display. Degrees refer to vector angles.

The purpose of the present study was to examine the effects of such a system interaction on the acquisition and transfer of skill. The major variable under study was the degree of P × R interaction present in the skill task (the amount of loss of lift resulting from lateral deflections of a control device in an aircraft simulator). Three hypotheses were tested concerning the effects of degree of component interaction on transfer performance levels: (*a*) the benefits of part-task practice, as measured on the whole task, will decrease as the degree of component interaction present in the whole task is increased; (*b*) proficiency levels during transfer will increase with the degree of similarity of component interaction in the training and the transfer systems; and (*c*) transfer from a higher to a lower degree of component interaction will result in a greater level of transfer performance than will transfer from a lower to a higher degree of interaction.

The first hypothesis merely acknowledges that training on individual components cannot provide *S* with practice on a coordination of his responses such as will be required in operation of the whole task when that whole task involves an interaction of components. The second and third hypotheses indicate the expectation of an asymmetrical generalization gradient, i.e., transfer from "difficult" to "easy" versions of the task is expected to involve a greater degree of transfer than that found with transfer from "easy" to "difficult" (**9, 12**). Thus, this experiment has two parts: Part I represents a test of the first hypothesis, while Part II tests the second and third. Data collection for both parts was accomplished concurrently.

METHOD

Apparatus

A version of the OSU Pilot Training Research Simulator (**4**) served as the skill task. The essential features of this device are illustrated in Fig. 1 and 2. The *S* was provided with a 5-in. cathode ray tube (CRT) display and a spring-centered, two-dimensional control column. He attempted to manipulate the control so as to hold a cursor at the center of a 1-in. circular reticle. Only the cursor, identified as an own; position indicator, moved about the display-the reticle was fixed at the center of the CRT

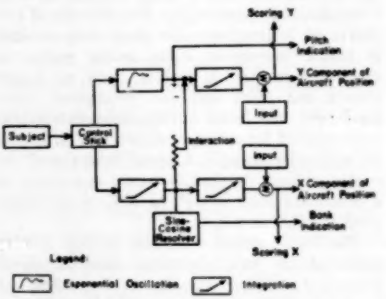


FIG. 2. The control dynamics of the *X* and *Y* (roll and pitch) dimensions of the tracking task.

and thus *S* was provided with a compensatory display of tracking information.

Right or left deflections of the control column generated a voltage *X* which was subjected to two stages of analog integration. These transformations provide for a simulation of azimuth corrections in an aircraft, and the output affected lateral movements of the own-position indicator on the CRT. Fore or aft control deflections generated a voltage *Y* which was transformed by an exponential oscillatory function plus one stage of analog integration, a combination which simulates the underdamped response of an aircraft to pitch corrections. The output of the first integral term in the *X* dimension represents bank angle, and this signal was fed to a sine-cosine resolver, the output of which determined the angle of incidence of the own-position indicator. Rotation of this indicator represented the lateral attitude of the "wings" of *S*'s aircraft.

The desired physical interaction was achieved by coupling the *X* to the *Y* dimensions of the control system as shown in Fig. 2; the voltage signal representing bank angle was subjected to attenuation, rectified, and summed with the *Y* signal at the integral transformation in the pitch control portion of the system. This rectified signal always appeared as a negative quantity at the summing point, and thus *S* was required to supply a positive signal of like magnitude in order to keep the cursor from moving off the lower portion of the CRT. This was achieved by deflecting the control column back from its spring-centered position. The attenuation enabled *E* to vary the amount of component interaction from zero to a maximum, at which level the cursor followed a vector path whose angle was 76° from horizontal. At zero interaction, lateral deflections of the control column affected only the *X* or roll dimension of the system, and thus *S* could respond to both dimensions independently. The presence of any degree of interaction other than zero resulted in cursor movement along vector paths, as illustrated in Fig. 1, and *S* could no longer correct the pitch and roll dimensions independently. Six levels of component interaction were selected for study, as shown in Fig. 1. At $(P \times R)_3$, for example, a lateral deflection of the control column resulted in cursor movement at a downward angle of 23°, or 30% of the maximum possible vector.

The input signals indicated in Fig. 2 were produced by two electronic random signal generators. These signals were peaked at 3 cpm with 95% of their power distributed below 12 cpm. The amplitude distribution of each input was normal with an *SD* of 3.9 v. The diameter of the 1-in. circular reticle corresponded to ±3 *SD*'s of the input signal. With these inputs the own-position indicator was forced away from the center of the reticle in a random manner, and *S* was instructed to compensate for his error by appropriate control deflections. The *X* input was statistically equivalent to, but independent of and uncorrelated with, the *Y* input.

Subjects and Procedure

A total of 160 Air Force ROTC students were paid for service in this experiment. None had had prior flight training. Each *S* served for four 45-min. sessions with 24 hr. between sessions. During the first session, all *S*s received a total of eight 30-sec. tracking trials following detailed instructions and demonstrations by *E*. There were a total of 16 30-sec. trials during each of the subsequent sessions, and practice was carried out in four-trial blocks. There was 30-sec. rest between trials within blocks and 4-min. rest between blocks. The *S*s served in pairs, alternately tracking and observing in four-trial blocks.

Part I.—Four degrees of P × R interaction were used in a test of the hypothesis that a part-task training schedule will be progressively less efficient as the component interaction present in the whole task is increased: $(P \times R)_0$, $(P \times R)_1$, $(P \times R)_3$, and $(P \times R)_9$. As indicated in Table 1, Groups 1 through 4 served as the experimental groups. There were totals of 12 30-sec. trials on pitch corrections alone and 12 trials on roll corrections alone for each *S* over Sessions 1 and 2. Transfer to the designated whole-task condition took place in Session 3. Groups 5, 6, 8, and 10 served as control groups against which to compare the performance levels of Groups 1–4, respectively, in Session 3.

Part II.—As a test of the second and third hypotheses six groups began and continued training on the two-dimensional tracking task where level of component interaction defined groups. Table 1 indicates that Groups 5, 6, 7, 9, and 10 served as experimental groups with each *S* receiving a total of 40 30-sec. trials of tracking under the interaction level of his group prior to transfer to $(P \times R)_3$ during Session 4. Thus, Group 8 provided the control procedure against which to measure levels of transfer. All groups received a total of 16 trials on $(P \times R)_3$ during the transfer session. As shown in Table 1, Groups 6 and 7 transferred to a higher degree of component interaction compared to that present in the training system, while Groups 9 and 10 transferred to a lower level, and Group 5 transferred from a system involving no component interaction. The

TABLE 1

EXPERIMENTAL DESIGN FOR PARTS I AND II
(16 *Ss* per group)

Part	Groups	Sessions			
		1	2	3	4
I	1	P, R[a]	P, R[b]	$(P \times R)_0$	$(P \times R)_0$
I	2	P, R	P, R	$(P \times R)_1$	$(P \times R)_1$
I	3	P, R	P, R	$(P \times R)_5$	$(P \times R)_5$
I	4	P, R	P, R	$(P \times R)_9$	$(P \times R)_9$
I, II	5	$(P \times R)_0$	$(P \times R)_0$	$(P \times R)_0$	$(P \times R)_5$
I, II	6	$(P \times R)_1$	$(P \times R)_1$	$(P \times R)_1$	$(P \times R)_5$
II	7	$(P \times R)_3$	$(P \times R)_3$	$(P \times R)_3$	$(P \times R)_5$
I, II	8	$(P \times R)_5$	$(P \times R)_5$	$(P \times R)_5$	$(P \times R)_5$
II	9	$(P \times R)_7$	$(P \times R)_7$	$(P \times R)_7$	$(P \times R)_5$
I, II	10	$(P \times R)_9$	$(P \times R)_9$	$(P \times R)_9$	$(P \times R)_5$

[a] Four trials pitch corrections alone, four trials roll corrections alone.
[b] Eight trials pitch corrections alone, eight trials roll corrections alone.

choice of interaction level $(P \times R)_5$ as the transfer condition was arbitrary in the sense that it is not necessarily typical of any particular skill system now in use.

Scoring

On each trial *E* recorded three measures of tracking performance: (*a*) two integrated error squared scores, one for the *X* and one for the *Y* dimension, and (*b*) one integrated cross-product of the *X* and *Y* errors. Although this was a two-dimensional tracking task, it was possible to derive a single, unitary score of tracking proficiency. This score, identified as the standard error (SE_{xy}), was determined by combining the above measures as follows:

$$[\int e^2_x + \int e^2_y - 2\int e_x e_y]^{\frac{1}{2}}$$

where e represents tracking error. The squaring, multiplying, and analog integration operations were performed continuously by electronic circuits during each 30-sec. trial. These SE_{xy} scores were later summed and averaged for each *S* over each four-trial block, a total of 2 min. of tracking.

RESULTS

Part I

Figure 3 summarizes the data in test of the hypothesis concerning the efficiency of part-task practice as a function of the degree of whole-task component interaction. During the part-task training trials, Blocks 1–6, the performance levels of Groups 1–4 are presented separately for the *X* (roll) and *Y* (pitch) components. The square roots of the integrated error squared scores (RMS_x and RMS_y) serve as the metrics here. The standard error (SE_{xy}) measure, discussed above, serves to describe the acquisition functions of the control groups on Sessions 1, 2, and 3 (Groups 5, 6, 8, and 10), and the same measure describes the transfer performance levels of Groups 1–4 during Sessions 3 and 4. Each point in Fig. 3 is the average performance level of the 16 *Ss* in each group over the four trials in each block. The interaction levels of the *X* and *Y* components in the whole or transfer tasks appear as the parameter in Fig. 3.

Training blocks.—Several points are apparent from visual inspection of the data defining tracking performance during Blocks 1–6. First, during the part-task training trials, none of the experimental groups show a significant change in performance as a function of trials. The RMS_x and RMS_y scores are uniformly low

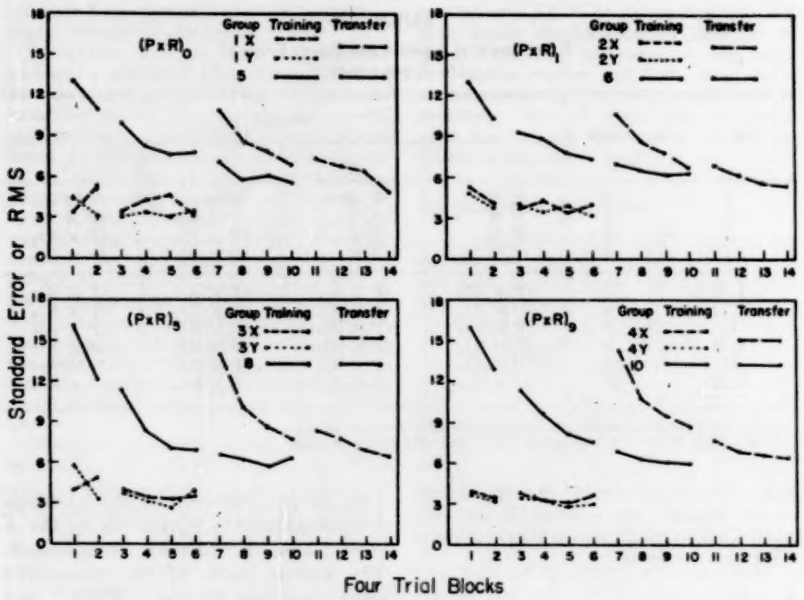


FIG. 3. Training and transfer performance levels for the control and experimental groups of Part I.

with little or no negative trend. It follows, then, that skill in tracking with the roll component alone and with the pitch component alone was easily acquired early in the training trials. This is in contrast to the four control groups (Groups 5, 6, 8, and 10) over the same period where initial tracking proficiency was quite low, especially for Groups 8 and 10. Further, skill levels on the whole task, as shown by the control groups, never did approach closely those levels achieved by Groups 1–4 during the training period. This, in part, may be a function of the differences in power of the inputs involved. The *RMS* value of the input for one-dimensional tracking was 3.9 v., while the standard error measure of the inputs for two-dimensional tracking was 5.5 v. These estimates were based on 10-min. samples of the input signals. Thus, Groups 1–4 did exceed the *RMS* level of the input (RMS_x and RMS_y on Block 6 averaged 3.3 v.), while the control groups never did achieve SE_{xy} measures of tracking performance lower than that for the input (5.8, 6.5, 6.3, and 6.0 v. for Groups 5, 6, 8, and 10, respectively, on Block 10).

Transfer blocks.—A comparison of the initial transfer performance levels of Groups 1–4, the Block 7 data of Fig. 3, with the Block 1 data of the appropriate control groups shows that some degree of positive transfer was achieved by each of the experimental groups. A convenient index to express this transfer is

$$\frac{C_1 - E_1}{C_1 - C_t} \times 100$$

where C_1 is the performance of a control group on Block 1 (13.1, 13.2, 16.2, and 16.3 v. for Groups 5, 6, 8, and 10, respectively), C_t is the tracking proficiency of the same groups on Block 7 (7.3, 7.1, 6.5, and 6.9 v.), and E_t is the performance level of the appropriate experimental group at Block 7 (11.4, 10.8, 14.5, and 14.6 v. for Groups 1–4, respectively). These indices are 29%, 39%, 18%, and 18% positive transfer for Groups 1–4, respectively.

These transfer indices substantiate, only in part, the hypothesis under test here. It was expected that the efficiency of pure part-task practice would be *progressively* lessened as the degree of component interaction present in the whole task was increased. The decrease in efficiency found here was not progressive in that the $(P \times R)_1$ level actually enjoyed a higher transfer index than did the $(P \times R)_0$ condition. However, both the higher levels of component interaction, Groups 3 and 4, did enjoy less benefit from part-task practice than did the lower degrees of interaction, Groups 1 and 2.

The above observations were borne out by the results of analyses of variance which were performed on the data of Session 3, Blocks 7–10. The results of these analyses show that differences between Groups 1 and 5 and between Groups 2 and 6 are statistically significant at $P < .05$,

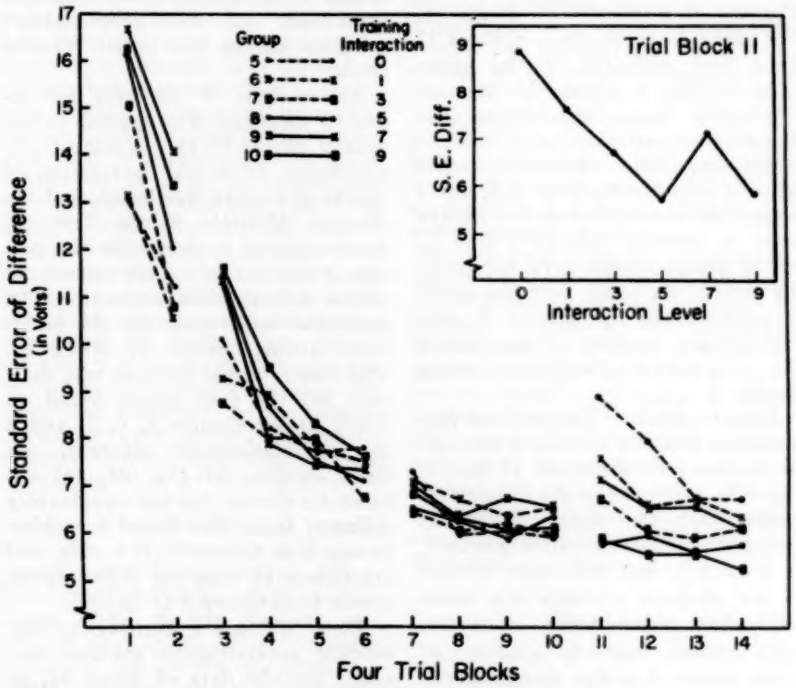


FIG. 4. Training and transfer performance levels for the control and experimental groups of Part II.

while those for Groups 3 and 8 and for Groups 4 and 10 are significant at $P < .001$.

Part II

Figure 4 summarizes the results of the test of the second and third hypotheses, concerning transfer performance levels as a function of similarity of component interaction in training and transfer systems. Again, the SE_{xy} metric was employed, and each point in Fig. 4 is the average performance level of 16 *Ss* in each group over the four trials in each block. As indicated earlier, there were 10 blocks of training trials prior to transfer to the $(P \times R)_5$ level of component interaction. Transfer was effected on Block 11. For convenience, the data of Block 11 have been replotted in the upper right of Fig. 4 where the training interaction level appears as the independent variable.

Training blocks.—As shown earlier by Fig. 3, there exist differences among the several interaction groups early in training (Blocks 1 and 2), but all groups appear to be approaching closely the same terminal levels of performance by Session 3 with considerable amounts of crossing of the acquisition functions during Session 2.

Transfer blocks.—The transfer performance levels of Groups 5 through 10, as shown by the Block 11 data of Fig. 4, in part confirm the hypotheses under test; the figure shows an asymmetrical generalization gradient, as predicted, but with that portion of the gradient covering the lower interaction levels being somewhat more regular than the gradient at levels higher than the control condition, $(P \times R)_5$. Further, the average transfer performance level of Groups 9 and 10 is somewhat better than the average of Groups 6 and 7, although this difference is not statistically significant ($P > .05$). Thus, transfer from higher to lower levels of component interaction and vice versa was in the direction predicted, but the difference was not reliable.

The same index as was used in Part I can serve to specify the amounts of transfer here. The value of C_1 is 16.2 v. and that of C_t is 5.7 v. (Group 8), while E_t is 8.9, 7.6, 6.7, 7.1, and 5.8 v. for Groups 5, 6, 7, 9, and 10, respectively. These indices as calculated are 70%, 82%, 90%, 87%, and 99% for Groups 5, 6, 7, 9, and 10. The trend of these percentages is, of course, in line with the above observations concerning the asymmetry of the generalization gradient during the initial transfer trials.

An analysis of variance for repeated measures was applied to the data of Blocks 11–14. A statistically significant ($P < .01$) interaction of Blocks × Groups was noted, and the Duncan Multiple Range Test (6) was employed to determine the pattern of significance among the several means defining that interaction. Of particular importance are the differences among means for Block 11. The results of the Duncan test show that (*a*) the four means found on Block 11 for Groups 5, 6, 7, and 8 are all statistically different, one from another, at $P < .05$, (*b*) the mean for Group 9 is not significantly different from that found for either Group 6 or Group 7 ($P > .05$), and (*c*) Group 10 does not differ significantly from Group 8 ($P > .05$).

Thus, whereas a statistically significant generalization gradient was noted for the data of Block 11, as predicted, the form of the gradient departs at one point from that

expected: the level of proficiency for Group 10 was significantly higher than anticipated. Coupled with the finding that transfer from "difficult" versions of the tracking task was not significantly greater than that from "easy" versions, it is concluded that the second hypothesis was confirmed only in part, while support for the third hypothesis is only suggested by the data.

Discussion

The above results indicate that, as a variable in complex skill acquisition and transfer, component interaction follows a fairly predictable pattern. Both the hypothesis concerning part-task practice (Part I) and the prediction of a generalization gradient (Part II) were supported by the data, with but a single departure from expectations in each case. The data, therefore, reveal no major departures from predictions, save the finding in Part II that transfer from higher to lower levels of component interaction was not significantly superior to transfer from lower to higher levels. A minor surprise comes from the rather low index of transfer found in Part I for Group 1: after a total of 24 trials of practice independently on the pitch and on the roll components, that group shows only 29% transfer to a whole task which involves no component interaction. As pointed out earlier, this version of the whole task does not require *S* to coordinate his responses to the two dimensions of the system, and as such it was expected that transfer from part to whole tasks would be high, at least in the order of 50% or more (**3, 5**).

The relatively low index of transfer found here for Group 1 can be contrasted with that reported by Briggs, Fitts, and Bahrick (**4**), where approximately 61% positive transfer was achieved by a group (Group 2) which transferred after 30 training trials to a system identical to that encountered here by Group 1. The major difference between the training tasks for these two groups is that the present Group 1 practiced the individual system components (*X* and *Y*) while the previous group practiced on a simplified but two-dimensional version of the common transfer task (that earlier training system required *S* to track through one rather than two stages of integration for bank maneuvers). It was shown in the earlier report that this reduction in number of integral lag terms reduced the complexity of both the perceptual and the motor responses required of *S*.

A critical consideration, then, is that the number of dimensions in a whole task be preserved in a training system to assure a reasonably high level of transfer. It follows from this that, in determining the characteristics of a training task, *simplification* rather than fractionation of a whole task may be expected to provide greater amounts of transfer, especially when the whole task requires or permits simultaneous responses to the several components or dimensions of the control system. Thus, it appears that reduction in an opportunity to integrate component skills, as in part practice, is far more detrimental to transfer performance than is simplification in either the perceptual or motor response requirements.

For the present, the above generalizations should be restricted to whole tasks in which simultaneity of responses is required or possible. The part-whole training data from tasks in which responses to the parts must be carried out in a serial order appear to be in conflict (**10, 11**). However, Seymour (**11**) suggests that in such tasks either pure part- or progressive part-task practice schedules are more efficient than the whole-task schedule, especially when the perceptual requirements of the whole task are complex.

Summary

Ten groups of 16 *S*s each received a total of 56 30-sec. trials tracking through simulated aircraft control dynamics. The variable under consideration was the degree of interaction

between two of the major components of this tracking system, the lateral and the vertical dimensions of flight control.

Three hypotheses were tested: (*a*) pure part-task practice will be progressively less beneficial, measured by transfer performance levels, as the degree of component (part) interaction is increased in the transfer or whole task, (*b*) transfer performance level will be proportional to the similarity of component interaction in the training and transfer tasks, with (*c*) transfer from higher to lower degrees of component interaction being greater than that found when transfer is from lower to higher degrees of such interaction.

The hypotheses were substantiated, in part, with the third hypothesis not receiving the expected support of statistical significance. The theoretical significance of relatively low indices of transfer found in the test of the first hypothesis was discussed. It was concluded that simplification rather than fractionation of a whole task provides for the more efficient training schedule when the whole requires or permits simultaneous responses to the several parts.

References

1. Bilodeau, E. A. A preliminary study of part-whole training on a complex compensatory pursuit task. *USAF, Personn. Train. Res. Cent. Lab. Note*, 1955, No. SCRL 55-7.
2. Bilodeau, E. A., & Bilodeau, I. McD. The contribution of component activities to the total psychomotor task. *J. exp. Psychol.*, 1954, **47**, 37–46.
3. Briggs, G. E., & Brogden, W. J. The effect of component practice on performance of a lever-positioning skill. *J. exp. Psychol.*, 1954, **48**, 375–380.
4. Briggs, G. E., Fitts, P. M., & Bahrick, H. P. Transfer effects from a single to a double integral tracking system. *J. exp. Psychol.*, 1958, **55**, 135–142.
5. Chambers, R. M. Task component training principles for modifying component and total task criterion performance. *Air Force Training and Personnel Symposium*, 1957, ARDC-TN (in press).
6. Duncan, D. B. A significance test for differences between ranked treatments in an analysis of variance. *Va. J. Sci.*, 1951, **2**, 171–189.
7. Ellson, D. G. The independence of tracking in two and three dimensions with the GE Pedestal Sight. *AAF, Air Materiel Command, Memo. Rep.*, 8 August 1947, No. TSEAA-694-2G. Pp. 6–16.
8. Gagné, R. M., & Foster, H. Transfer of training from practice on components in a motor skill. *J. exp. Psychol.*, 1949, **39**, 57–68.
9. Gibbs, C. B. Transfer of training and skill assumptions in tracking tasks. *Quart. J. exp. Psychol.*, 1951, **3**, 99–110.
10. McGuigan, F. J., & MacCaslin, E. F. Whole and part methods in learning a perceptual motor skill. *Amer. J. Psychol.*, 1955, **68**, 658–661.
11. Seymour, W. D. Experiments on the acquisition of industrial skills (Part 3). *Occup. Psychol.*, 1956, **30**, 94–104.
12. Szafran, J., & Welford, A. T. On the relation between transfer and difficulty of initial task. *Quart. J. exp. Psychol.*, 1950, **2**, 88–94.

(Received February 5, 1958)

Journal of Experimental Psychology
Vol. 56, No. 6, 1958

ACQUISITION OF A VERBAL MAZE AS A FUNCTION OF METHOD OF CORRECTION AND NUMBER OF ALTERNATE CHOICES PER UNIT [1]

R. F. THOMPSON AND W. J. BROGDEN

University of Wisconsin

Experiments by Brogden and Schmidt (**1, 2**) have shown that difficulty in acquisition of verbal mazes increases linearly as a function of the number of alternate choices per unit. These results hold for total time and total errors to the criterion, for mazes of 16 and 24 units, and for number of alternate choices per unit of 2, 4, 6, and 8 for the 16-unit maze and of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12 for the 24-unit maze. The variable number of alternate choices per unit is representative of the dimension of amount of discovery as defined by Melton (**4**). The increase in amount of discovery and the decrease in number of like responses that go with increase in number of alternate choices per unit probably account for the linear increase in difficulty of acquisition.

In view of the discovery component of number of alternate choices per unit, the acquisition of verbal mazes should be less difficult if discovery time is reduced by alteration of the correction procedure. This hypothesis was tested and confirmed by Thompson (**6**) by comparing the standard correction procedure with a modified correction procedure that allowed *S* to make only a single error at a choice point, following which *E* gave him the correct response. Thompson used mazes of 12 units in length that involved 2, 3, 4, 5, or 6 alternate choices per unit. The correct response for each unit was presented to *S* by a memory drum after he had spoken the correct response (correction procedure) or after a single response (modified correction procedure). This differed from the procedure of Brogden and Schmidt in which *S* spoke the alternate choices for each choice point until the correct response was made, when *E* spoke the word "Correct."

The present paper is a report of two experiments: one a methodological study that has led to modification of procedure, and the other designed to extend knowledge about the method of correction in acquisition of verbal mazes of greater length and with a greater number of alternate choices per unit than studied by Thompson (**6**).

PROCEDURE

Experiment I.—This experiment is comparable to the experiment reported by Brogden and Schmidt (**1**) except that the mazes were constructed from 2-digit numbers instead of words representing spatial locus and that the response was written instead of spoken. The mazes were 24 units in length with 2, 4, 6, 8, 10, and 12 alternate choices per unit.

The *S*'s responses were written on a Telautograph apparatus, the receiving machine being viewed in another room by *E*. Communication between *E* and *S* was by an intercommunication system. The *S* was given the choices for his maze, and instructed in the correction procedure. When *E* said "Begin," *S* was to write down the numbers used in any order until he wrote the correct one, whereupon *E* said "Forward" and *S* moved to the next unit, and so on, until one perfect repetition was completed by *S*. A 30-sec. intertrial interval was maintained throughout practice.

[1] This research was supported in part by a grant from the Research Committee of the Graduate School from funds provided by the Wisconsin Alumni Research Foundation.

Experiment II.—Mazes of 24 units length with 2, 4, 6, 8, 10, and 12 alternate choices per unit were learned to the criterion of one perfect trial by the standard correction procedure and by the modified correction procedure. There were 14 *Ss* in each of the 12 subgroups, each *S* serving in a single subgroup only. The *Ss* were volunteers from the classes in elementary psychology at the University of Wisconsin. The maze patterns were identical with those used in the above study employing Telautograph apparatus, and so identical with the Brogden and Schmidt (1) patterns, except that 2-digit numbers were coded for words.

The instructions and practice on the mazes followed the condition used by Brogden and Schmidt (1) except for use of the modified correction procedure and use of numbers instead of words. Thus, in the correction procedure, *S* was presented with the numbers to be used in his maze and told to respond verbally with the numbers until he gave the correct response, whereupon *E* said "Forward," and so on until one perfect repetition of the maze was completed. In the modified correction procedure, *S* was told to respond only once, whereupon *E* gave the correct response. In addition, no attempt was made in the present experiment to maintain the analogy with a spatial maze. The two *Es* each tested half the *Ss* in each experimental condition.[2] The *E* recorded each error and the time of the trial on prepared data sheets. A 30-sec. intertrial interval was maintained throughout practice.

Results

Experiment I.—Analyses of variance were completed on the time and total errors to criterion data. Both measures were found to increase significantly (5% level) with increasing number of choices per unit (for time, F equaled 6.89 with 5 and 54 *df*; for errors, F equaled 17.08 with 5 and 54 *df*). Tests for deviations from linearity (**5**, pp. 317–320) for both measures were not significant. Best-fitting straight lines were obtained for each measure by the method of least squares. The comparable data from the Brogden and Schmidt study (1) for 2, 4, 6, 8, 10, and 12 choices per unit was combined with the present data for both time and total errors measures and the pooled data treated statistically as a single experiment, with the two methods of running *S* as one variable. Analyses of variance were completed on both pooled measures, and for both, methods and mazes were significant at the 5% level, but the interactions of methods and mazes were in neither case significant. For both measures, the mean scores obtained in the present study are larger than those of the Brogden and Schmidt study.

The significance of the difference in slope of the best-fitting straight lines for the Brogden and Schmidt data and the present data was tested by the procedure described by Wilks (**7**, pp. 157–159) for both time and total errors. In both cases the t ratios were less than one, demonstrating no significant difference in the slopes of the straight lines obtained in the two studies. Hence, it may be concluded that the two different methods of running *Ss* yield functions of the same slope but of different absolute magnitude.

Experiment II.—Analyses of variance were completed on the data for time, total errors, total first errors, and trials to criterion and the results are given in Table 1. Since Bartlett tests for homogeneity of variance of the data indicated significant heterogeneity of variance, the 1% level of significance will be employed in the place of the 5% level. The results of the analyses show increase for all measures with increasing number of choices per unit (mazes). The two procedures differ in terms of time and total errors (larger scores for the correction procedure) but do not differ in terms of first errors and trials. The *E* is a significant variable for total errors

[2] The authors wish to express appreciation to Davis Chambliss who assisted in the conduct of the experiment by testing half of the *Ss*.

TABLE 1

ANALYSES OF VARIANCE

Source	df	Time		Total Errors		First Errors		Trials	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Mazes (M)	5	10,838,670	17.14*	1,128,100	31.51*	84,086	8.24*	282	3.17*
Procedures (P)	1	31,835,926	50.35*	7,716,429	215.53*	17,856	1.75	363	4.08
Experimenter (E)	1	3,329,018	5.26	603,241	16.85*	52,223	5.12	518	5.82
M × P	5	2,042,059	3.23*	623,841	17.42	6,653	.65	39	.44
M × E	5	416,364	.66	54,071	1.51	2,636	.26	50	.56
P × E	1	7,406,677	11.71*	474,991	13.27*	19,891	1.95	82	.92
M × P × E	5	1,252,134	1.98	124,922	3.49*	15,097	1.48	178	2.00
Error	144	632,341		35,801		10,209		89	

* $P = .01$.

only. The interaction of mazes and procedures is significant for time and total errors only. For both time and total errors, the procedure by *E*s interaction is significant, and nnally, the triple interaction of mazes by procedures by *E*s is significant for total errors only.

An orthogonal polynomials analysis was completed on all four measures of acquisition (**3**). For time and total errors for both the correction and modified correction procedures, the linear component, and only the linear component, is significant. The equations for the best-fitting straight lines were computed and are presented in Fig. 1 with plots of the curves and the empirical data points. For the trials measure under the modified correction procedure, no components were significant; under the correction procedure, the only significant component was the cubic. No components were significant for the first errors measure under the modified correction procedure. For the first errors measure under the correction procedure, however, the linear, quadratic, and cubic components were statistically significant. Equations for the trials and first error measures under the correction procedure were computed and are presented in Fig. 2 with plots of the curves and the empirical data points.

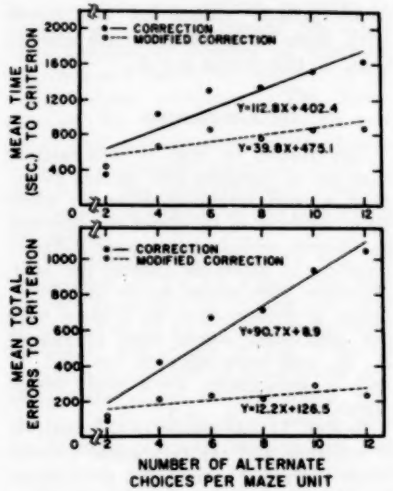


FIG. 1. Relation between time (above) and total error (below), and number of alternate choices per unit with correction and modified correction procedures as parameter.

The significant effect of *E* and the interaction of mazes, procedures, and *E* on the total errors measure (Table 1) were at first thought to be due to sampling differences, since the two *E*s ran *S*s in different semesters, thus failing to fulfill completely the requirement of randomization. In or-

der to check on sampling differences, the *Ss'* grades and ACE scores were obtained and analyses of variance were completed to compare the two groups of *Ss* on these measures. No significant differences between groups were found. Hence it seems reasonable to conclude that the differences are a function of *E* and are not due to sampling error. Total errors to the criterion are plotted against number of alternate choices per unit in Fig. 3 with *Es* and procedures as the parameters.

The curves for both *Es* are essentially the same with the modified correction procedure, but with the correction procedure the curves of E_2 for both time and errors show much greater values for the mazes with more than six alternate choices than do the curves for E_1.

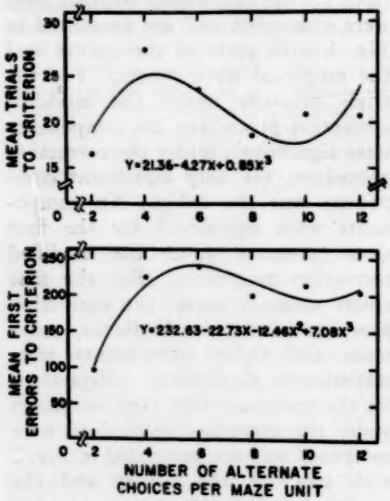


FIG. 2. Relation between trials to criterion, correction procedure (above); and relation between first errors to criterion, correction procedure (below), and number of alternate choices. In the equation for each of the two curves, $x = 0$ is the midpoint (7) for the scale of the independent variable, number of alternate choices per maze unit. Each positive or negative unit of x is equal to two units of the independent variable.

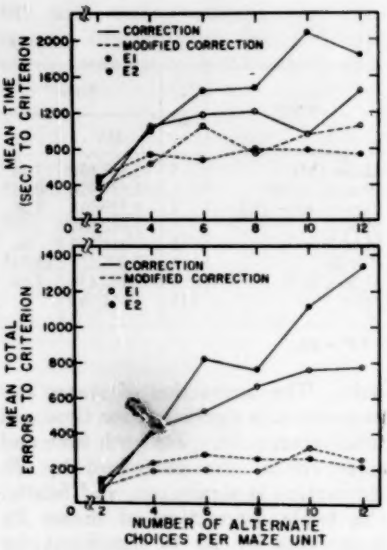


FIG. 3. Relation between time (above) and total errors (below) and number of alternate choices with *E* and Procedure as parameters.

DISCUSSION

The results of Exp. I demonstrate that use of numbers and written rather than spoken responses have no effect upon the form of the functions obtained in 24-unit verbal mazes over differing numbers of alternate choices per unit. Numbers were found to be simpler in terms of presentation to *S* than were the directional words used previously (**1, 2**), and were adopted for Exp. II and in subsequent work with verbal mazes.

In Exp. II, the results of the analyses of variance and the orthogonal polynomials analyses for time and total errors are in complete agreement with those obtained by Thompson (**6**) for the 12-unit directional word verbal mazes. Both measures increase linearly with

increasing numbers of alternate choices per unit, but the slopes of the functions for the correction procedure are steeper than those for the modified correction procedure (see Table 1 and Fig. 1). For the measures trials and first errors, no effect of number of choices was obtained under the modified correction procedure. With the correction procedure, the effect on trials has only significant cubic components whereas that for first errors has significant linear, quadratic and cubic functions (see Fig. 2). These results are similar to those obtained by Thompson (6) in that significant effects for trials and first errors were obtained only for the correction procedure, but different in that the functions were linear. The difference between 12 and 24 units of length for the mazes in the two experiments probably accounts for the difference in results. The 24-unit maze represents a learning task of considerably greater difficulty. The questionable validity of trials in comparing mazes (see discussion by Brogden and Schmidt [1], and Thompson [6]) and the lack of any reasonable hypothesis to account for the cyclical variation in number of trials to acquire the maze as a function of number of alternate choices place the fitted function in an indeterminate category. The fitted function for the relationship of first errors and number of alternate choices with the correction procedure must be placed in the indeterminate category also because of the lack of any reasonable hypothesis.

The significant effects of *E*s and the interactions of *E*s by Procedures and *E*s by Procedures by Mazes for total errors and of *E*s by Procedures for time (Table 1) are shown in Fig. 3 to be primarily a function of more errors and more time under the correction procedure for *S*s run by E_2 on mazes with six or more alternate choices than for *S*s tested by E_1. That interaction occurs between *E* and *S* in experiments on learning is self-evident, but in general *E* has not been considered to be a critical variable in determining the experimental results. In the present experiment, there is much more interaction between *E* and *S* with the correction procedure than with the modified correction procedure. The quickness of *E*'s response "Forward" when *S* makes the correct response, his tone of voice, and other personal characteristics appear to be representative of the kinds of difference in behavior between *E*s that might account for the difference in performance by *S*s. Even though precise information is lacking on what differences between *E*s are responsible for the differences in experimental results, the effect is of considerable importance. Experimenter effects may be responsible for differences in experimental results when the same experiment has been replicated in different laboratories by different *E*s. In any case, consideration must be given always to the possibility in many experiments on learning that the obtained results are in part a function of the personal characteristics of *E*.

Summary

When the experimental procedure of Brogden and Schmidt (1, 2) was changed in Exp. I from a vocal to a written response and from words to 2-digit numbers, the relations between time and total errors to the criterion and number of alternate choices per maze unit were comparable except for the absolute magnitude of the measures.

In Exp. II *S*s learned the mazes of Exp. I under two procedures—the standard correction procedure and a modified correction procedure. Two *E*s tested half the *S*s in each of 12 conditions (number of alternate choices per unit of 2, 4, 6, 8, 10, or 12, and correction and modified correction procedure).

Significant differences in slope were found for the linear functions of time and of total errors and number of alternate choices for the correction and modified correction procedures comparable to those obtained by Thompson (6) for the 12-unit word maze. Statistically significant effects of *E*s were obtained for the functions of time and total errors primarily under the correction procedure. This latter result suggests that consideration must be given to the possibility in many experiments on learning that the results may in part be a function of the personal characteristics of *E*.

References

1. Brogden, W. J., & Schmidt, R. E. Acquisition of a 24-unit verbal maze as a function of number of alternate choices per unit. *J. exp. Psychol.*, 1954, **48**, 335–338.
2. Brogden, W. J., & Schmidt, R. E. Effect of number of choices per unit of a verbal maze on learning and serial position errors. *J. exp. Psychol.*, 1954, **47**, 235–240.
3. Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research.* New York: Hafner, 1953.
4. Melton, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* (Rev. ed.) New York: Macmillan, 1950. Pp. 668–690.
5. Snedecor, G. W. *Statistical methods.* (3rd ed.) Ames: Iowa State Coll. Press, 1940.
6. Thompson, R. F. A comparison of correction and modified correction procedures on the acquisition of a 12-unit verbal maze. *J. exp. Psychol.*, 1958, **56**, 443–447.
7. Wilks, S. S. *Mathematical statistics.* Princeton: Princeton Univer. Press, 1950.

(Received February 10, 1958)

Journal of Experimental Psychology
Vol. 56, No. 6, 1958

MOTIVATION SHIFT IN A COMPLEX LEARNING TASK [1]

DAVID BIRCH

University of Michigan

Theoretical analyses and experimental investigations growing out of the Hullian postulate relating habit and drive, $R = f(E) = f(H \times D)$, have emphasized the importance of the initial response hierarchy in determining performance (**15, 18**). According to the theory, an increase in the strength of D should result in improved performance if the response hierarchy is such that the correct response tendency is dominant, but may result in no difference or even in poorer performance if incorrect competing responses are relatively equal in strength to the correct response (a so-called complex learning task). Experimental studies, which have produced data relevant to the theory, have employed two methods for obtaining differences in D among groups. In some studies (**3, 6, 7, 11, 12, 16, 17, 18, 19, 20**), *S*s were selected for degrees of emotionality by the Taylor scale (**21**). Other experiments (**1, 9, 13**) have performed experimental manipulations designed to produce differential D levels in *S*s while combinations of selection and manipulation have also been used (**5, 8, 10, 14**). In these studies, which range considerably in complexity, the assessment of the initial response hierarchy was made on a priori grounds, and traditional learning measures, such as trials and/or errors to criterion, were used to evaluate the theory.

The present study is concerned with the effects of an experimentally introduced shift in motivation on performance in a complex task but utilizes an alternative method. In brief, the application of the method is as follows: *S*s are allowed the opportunity to learn R_1 to two sets of single stimuli, S_{A1} and S_{B1}, and R_2 to two other sets of single stimuli, S_{A2} and S_{B2}, and are tested on the single sets, S_{A1}, S_{A2}, S_{B1}, and S_{B2}, and also on the compound sets, $S_{A1}S_{B1}$ and $S_{A2}S_{B2}$. At a given level of training *S*s are divided into two groups, one of which received instructions designed to increase its level of D over that of the other, and tested again on the complete sets of stimuli. Thus, the method attempts an assessment of the response hierarchy of the two groups before and after a differential shift in motivation. A model for response tendency combination (**4**), constructed to treat the problem of the probability of correct response to a two-element compound stimulus as a function of the probability of correct response to the two elements where two responses are available to *S*, is appropriate for the assessments.

Application of the model allows the determination of the values of four parameters from the data. These parameters are $\Pr(D_A \geq d)$, $\Pr(D_A \leq -d)$, $\Pr(D_B \geq d)$, and $\Pr(D_B \leq -d)$ where D_A and D_B are the differences between the correct and incorrect E's for Stimuli A and B, respectively. The parameter, d, is chosen so that the following conditional probabilities are satisfied:

$$\Pr(A,+|D_A \geq d) = 1, \quad \Pr(A,+|D_A \leq -d) = 0, \quad \Pr(A,+|-d < D_A < d) = .5,$$
$$\Pr(B,+|D_B \geq d) = 1, \quad \Pr(B,+|D_B \leq -d) = 0, \quad \text{and} \quad \Pr(B,+|-d < D_B < d) = .5.$$

[1] This study was carried out under a project with the Michigan Bell Telephone Company. Daniel Solomon and Astrid Beigel served as *E*s.

For convenience, a single symbol, d, has been used in these expressions though there is no necessity that $|-d| = d$ or that the d's be equal for A and B.

It follows that

$$\Pr(A,+) = (.5)[1+\Pr(D_A \geq d)-\Pr(D_A \leq -d)] \quad [1]$$

and

$$\Pr(B,+) = (.5)[1+\Pr(D_B \geq d)-\Pr(D_B \leq -d)] \quad [2]$$

where $\Pr(A,+)$ and $\Pr(B,+)$ are the probabilities of correct response to the single stimulus sets, A and B. Four classes of compound stimuli, (A,+;B,+), (A,+;B,−), (A,−;B,+) and (A,−;B,−), in the expected proportions, $\Pr(A,+)\cdot\Pr(B,+)$, $\Pr(A,+)\cdot\Pr(B,-)$, $\Pr(A,-)\cdot\Pr(B,+)$ and $\Pr(A,-)\cdot\Pr(B,-)$, where $\Pr(A,-)=1-\Pr(A,+)$ and $\Pr(B,-) = 1-\Pr(B,+)$, result from the construction of the compound stimuli. The probabilities of correct and incorrect response to these four classes of the compound stimuli are:

$$\Pr(AB,+|A,+;B,+)[\Pr(A,+)\cdot\Pr(B,+)] = (.5)[\Pr(A,+)\cdot\Pr(B,+)+\Pr(D_A \geq d)\cdot\Pr(B,+) + \Pr(D_B \geq d)\cdot\Pr(A,+)-\Pr(D_A \geq d)\cdot\Pr(D_B \geq d)] \quad [3]$$

$$\Pr(AB,-|A,+;B,+)[\Pr(A,+)\cdot\Pr(B,+)] = \Pr(A,+)\cdot\Pr(B,+)[1-\Pr(AB,+|A,+;B,+)] \quad [4]$$

$$\Pr(AB,+|A,+;B,-)[\Pr(A,+)\cdot\Pr(B,-)] = (.5)[\Pr(A,+)\cdot\Pr(B,+)-\Pr(D_A \geq d)\cdot\Pr(B,+) - \Pr(D_B \geq d)\cdot\Pr(A,+)+\Pr(D_A \geq d)] \quad [5]$$

$$\Pr(AB,-|A,+;B,-)[\Pr(A,+)\cdot\Pr(B,-)] = \Pr(A,+)\cdot\Pr(B,-)[1-\Pr(AB,+|A,+;B,-)] \quad [6]$$

$$\Pr(AB,+|A,-;B,+)[\Pr(A,-)\cdot\Pr(B,+)] = (.5)[\Pr(A,+)\cdot\Pr(B,+)-\Pr(D_A \geq d)\cdot\Pr(B,+) - \Pr(D_B \geq d)\cdot\Pr(A,+)+\Pr(D_B \geq d)] \quad [7]$$

$$\Pr(AB,-|A,-;B,+)[\Pr(A,-)\cdot\Pr(B,+)] = \Pr(A,-)\cdot\Pr(B,+)[1-\Pr(AB,+|A,-;B,+)] \quad [8]$$

$$\Pr(AB,+|A,-;B,-)[\Pr(A,-)\cdot\Pr(B,-)] = (.5)[\Pr(A,+)\cdot\Pr(B,+)-\Pr(D_A \geq d)\cdot\Pr(B,+) - \Pr(D_B \geq d)\cdot\Pr(A,+)+\Pr(D_A \geq d)\cdot\Pr(D_B \geq d)] \quad [9]$$

and

$$\Pr(AB,-|A,-;B,-)[\Pr(A,-)\cdot\Pr(B,-)] = \Pr(A,-)\cdot\Pr(B,-)[1-\Pr(AB,+|A,-;B,-)] \quad [10]$$

In addition, the total probability of correct response to the compound stimuli, arrived at by summing Equations 3, 5, 7, and 9, is

$$\Pr(AB,+) = 2\Pr(A,+)\cdot\Pr(B,+)+\Pr(D_A \geq d)[(.5)-\Pr(B,+)] + \Pr(D_B \geq d)[(.5)-\Pr(A,+)] \quad [11]$$

The purpose of the present experiment is to investigate the effects of a shift in motivation on performance in the light of the model. Consider Stimulus A where $\Pr(D_A \geq d)$ and $\Pr(D_A \leq -d)$ have known values under the standard motivational conditions of the experiment. According to a multiplicative relationship between H and D, an increase in D should result in an increase in $|D_A|$ and tend to increase both $\Pr(D_A \geq d)$ and $\Pr(D_A \leq -d)$. Conversely, a decrease in D would be expected to decrease both parameters. Corresponding effects should occur for Stimulus B.

An increase in D has the following implications for Equations 1–11: $\Pr(A,+)$ and $\Pr(B,+)$ involve compensating effects and should not change appreciably; $\Pr(AB,+|A,+;B,+)[\Pr(A,+)\cdot\Pr(B,+)]$ and $\Pr(AB,-|A,-;B,-)[\Pr(A,-)\cdot\Pr(B,-)]$ should increase; $\Pr(AB,-|A,+;B,+)[\Pr(A,+)\cdot\Pr(B,+)]$ and $\Pr(AB,+|A,-;B,-)[\Pr(A,-)\cdot\Pr(B,-)]$ should decrease while $\Pr(AB,+|A,+;B,-)[\Pr(A,+)\cdot\Pr(B,-)]$, $\Pr(AB,-|A,+;B,-)[\Pr(A,+)\cdot\Pr(B,-)]$, $\Pr(AB,+|A,-;B,+)[\Pr(A,-)\cdot\Pr(B,+)]$, $\Pr(AB,-|A,-;B,+)[\Pr(A,-)\cdot\Pr(B,+)]$, and $\Pr(AB,+)$ are indeterminate because of a lack of knowledge of the relative magnitude of the effects. The experiment is designed to test the predicted shifts in the theoretical parameters, $\Pr(D_A \geq d)$, $\Pr(D_A \leq -d)$, $\Pr(D_B \geq d)$, and $\Pr(D_B \leq -d)$ and in the observed values of the four relevant categories of the compound stimuli.

Method

Subjects.—The *Ss* were 108 males obtained from the introductory psychology class and paid at the rate of $1.00 per hour for participating in the experiment. Twelve *Ss* were discarded for failure to follow instructions.

Apparatus.—In order to provide learning, two sets of stimuli, S_A = letter pairs (L) and S_B = number pairs (N), of 20 items each were

presented individually in a random sequence on flash cards for 10 sec. with the correct response exposed during the first 5 sec. One-half of the letter pairs (S_{A1}) and one-half of the number pairs (S_{B1}) were to be associated with the response, JIX, and the other half of each set (S_{A2}) and (S_{B2}) were to be associated with the response, DAC. The 20 letter pairs were constructed by pairing the 10 letters, B, F, G, H, K, N, Q, S, Y, and Z, in such a fashion that each letter appeared in the first position twice, once to be associated with JIX and once with DAC, and in the second position twice, again once to be associated with each response. No two letters were ever paired with each other more than once. The 10 digits, 0 through 9, were treated in like manner.

Three sets of compound stimuli (LN) were constructed for use in the test series by joining on one flash card a letter pair and a number pair, chosen at random without replacement under the restriction that the same response be correct for both the letter pair and the number pair. Table 1 shows the complete set of stimuli and the assignment to the two response categories.

Procedure.—Twelve groups, composed of approximately nine *Ss* each, participated as units in the experiment. The *Ss* were introduced to the apparatus and instructed that they were to learn the correct response, JIX or DAC, to each of the 40 stimuli. Learning series were alternated with test series, and *Ss* recorded their responses on prepared sheets of paper only during the test series when the stimuli alone were presented. Two learning and three test series were given. In the first learning series, each of the 40 stimuli along with the correct response was shown once while the second learning series allowed two trials on each of the 40 S-R combinations.

Each test series was made up of the 40 single stimulus pairs and 20 compound stimuli as indicated in Table 1. After the second test all *Ss* received a third test without an intervening learning series. Between the second and third tests *Ss* in the Control group ($N = 44$) were told only that they would receive another test immediately whereas *Ss* in the Increased Motivation (IM) group ($N = 52$) were instructed that they were to be paired at random with each other and that an extra dollar would be awarded to that member of each pair who scored higher on the next test, to be given immediately. These instructions for the IM group were adapted from those used by Atkinson (2).

Results

The differences in performance for the IM and Control groups on Tests 2 and 3 relate directly to the hypothesis under test. In order to gain more information, both about the model and about the effect of the motivation shift, each group was divided at its median number of compound stimuli correct on Test 2 to form a High and a Low subgroup for the two conditions. Equal-sized subgroups were obtained by assigning at random two

TABLE 1

Assignment of the Letter Pairs and Number Pairs to the Two Response Categories, JIX and DAC, and the Compound Stimuli Used in the Three Tests

Letter Pairs (L)		Number Pairs (N)		Compound Stimuli (LN)					
				Test 1		Test 2		Test 3	
JIX	DAC	JIX	DAC	JIX	DAC	JIX	DAC	JIX	DAC
BF	BN	09	06	BF 12	BN 63	BF 09	BN 92	BF 45	BN 59
FG	FQ	12	17	FG 27	FQ 17	FG 68	FQ 80	FG 12	FQ 75
GZ	GS	27	24	GZ 50	GS 41	GZ 45	GS 59	GZ 09	GS 06
HQ	HY	31	38	HQ 84	HY 92	HQ 31	HY 38	HQ 68	HY 24
KB	KZ	45	41	KB 45	KZ 06	KB 27	KZ 63	KB 84	KZ 80
NH	NF	50	59	NH 96	NF 75	NH 12	NF 17	NH 73	NF 41
QS	QG	68	63	QS 31	QG 38	QS 73	QG 75	QS 96	QG 92
SK	SH	73	75	SK 73	SH 59	SK 50	SH 24	SK 27	SH 17
YN	YK	84	80	YN 09	YK 24	YN 84	YK 41	YN 50	YK 38
ZY	ZB	96	92	ZY 68	ZB 80	ZY 96	ZB 06	ZY 31	ZB 63

TABLE 2

OBSERVED PROPORTION OF CORRECT AND INCORRECT RESPONSES BY THE TWO LEVELS OF THE INCREASED MOTIVATION AND CONTROL GROUPS TO THE LETTERS, NUMBERS, AND COMPOUND STIMULI FOR TESTS 2 AND 3

Responses	Increased Motivation				Control			
	High		Low		High		Low	
	Test 2	Test 3	Test 2	Test 3	Test 2	Test 3	Test 2	Test 3
Pr(L, +)	.658	.692	.571	.527	.739	.668	.593	.577
Pr(N, +)	.717	.758	.550	.554	.741	.759	.580	.561
Pr(LN, +)	.802	.756	.537	.575	.827	.786	.577	.645
Pr(LN, +|L, +; N, +) [Pr(L, +)·Pr(N, +)]	.442	.500	.221	.229	.527	.459	.261	.286
Pr(LN, −|L, +; N, +) [Pr(L, +)·Pr(N, +)]	.046	.044	.088	.060	.041	.048	.075	.027
Pr(LN, +|L, +; N, −) [Pr(L, +)·Pr(N, −)]	.119	.071	.112	.112	.130	.118	.127	.143
Pr(LN, −|L, +; N, −) [Pr(L, +)·Pr(N, −)]	.050	.077	.150	.127	.041	.043	.130	.120
Pr(LN, +|L, −; N, +) [Pr(L, −)·Pr(N, +)]	.185	.160	.133	.158	.141	.184	.148	.175
Pr(LN, −|L, −; N, +) [Pr(L, −)·Pr(N, +)]	.044	.054	.108	.108	.032	.068	.095	.073
Pr(LN, +|L, −; N, −) [Pr(L, −)·Pr(N, −)]	.056	.025	.071	.077	.030	.025	.041	.041
Pr(LN, −|L, −; N, −) [Pr(L, −)·Pr(N, −)]	.058	.069	.117	.131	.059	.055	.123	.134

Ss in the IM group and one *S* in the Control group with scores of 14 (the median scores for the two groups) to the Low subgroups. Table 2 contains the observed proportion of correct and incorrect responses by the two levels of the IM and Control groups on the Letters, Numbers, Compound Total, and the four classes of the compound stimuli for Tests 2 and 3. The bases for the proportions are 520 (26 *Ss* × 20R's) for the IM subgroups and 440 (22 *Ss* × 20R's) for the Control subgroups.

The patterns of change from Test 2 to Test 3 over the four subgroups for the Letters, Numbers, and Compound Total are irregular. More important to the test of the theory, however, are the changes in the responses for the four relevant categories of the compound stimuli. Table 3 presents the Test 3 minus Test 2 differences for the two levels of the IM and Control groups for these categories classified in terms of the (L,+;N,+) and (L,−;N,−) stimuli and the response direction to LN same as, or opposite to, that of the elements. According to the theory it is anticipated that the differences for the IM group should be positive for

$$(LN,+|L,+;N,+) \qquad [Pr(L,+)\cdot Pr(N,+)]$$

and

$$(LN,-|L,-;N,-) \qquad [Pr(L,-)\cdot Pr(N,-)],$$

and negative for

$$(LN,-|L,+;N,+) \qquad [Pr(L,+)\cdot Pr(N,+)]$$

and

$$(LN,+|L,-;N,-) \qquad [Pr(L,-)\cdot Pr(N,-)].$$

That this general pattern appears to hold for the IM group and a somewhat different pattern for the Control group may be seen in Table 3.

Table 4 summarizes the analysis of variance for the four dimensional factorial design of Table 3. Four Test 3 minus Test 2 difference scores, one for each category of concern, were determined for each of the 96 *Ss*, and the differences for categories

$$(LN,-|L,+;N,+) \qquad [Pr(L,+)\cdot Pr(N,+)]$$

and

$$(LN,+|L,-;N,-) \qquad [Pr(L,-)\cdot Pr(N,-)]$$

were multiplied by -1 so that the positive scores would indicate the direction of the hypothesis.

The main effect for Motivation Conditions fails to reach the 5% level of significance ($F = 3.15$, 1 and 92 *df*, $P \leq .10$), but the MC × L interaction is significant at the 1% level.

TABLE 3

OBSERVED TEST 3 MINUS TEST 2 PROPORTION DIFFERENCES FOR THE TWO LEVELS OF THE INCREASED MOTIVATION AND CONTROL GROUPS ON THE FOUR SELECTED CATEGORIES OF TABLE 2, CLASSIFIED IN TERMS OF STIMULI, (L,+;N,+) AND (L,−;N,−), AND RESPONSES TO LN, SAME OR OPPOSITE TO THE ELEMENTS OF THE STIMULI

Motiv.	Stimuli	Increased Motivation		Control	
		Response to LN		Response to LN	
		Same	Opposite	Same	Opposite
High	(L,+;N,+)	.058	−.002	−.068	.007
	(L,−;N,−)	.011	−.031	−.004	−.005
Low	(L,+;N,+)	.008	−.028	.025	−.048
	(L,−;N,−)	.014	.006	.011	.000

TABLE 4

ANALYSIS OF VARIANCE ON THE TEST 3 MINUS TEST 2 DIFFERENCES FOR THE FOUR SELECTED CATEGORIES OF THE COMPOUND STIMULI

Source	*df*	*F*
Between *Ss*	95	
Motivation Cond. (MC)	1	3.15
Levels (L)	1	1.17
MC × L	1	7.89**
Error (b)	92	(3.55)
Within *Ss*	288	
Stimuli (S)	1	0.31
Responses (R)	1	0.31
MC × S	1	0.78
MC × R	1	2.95
L × S	1	4.14*
L × R	1	0.04
S × R	1	0.38
MC × L × S	1	3.38
MC × L × R	1	2.07
MC × S × R	1	3.18
L × S × R	1	2.57
MC × L × S × R	1	3.88*
Error (w)	276	(2.74)

Note.—Mean squares for error in parentheses.
* $P = .05$.
** $P = .01$.

Because of the significant MC × L interaction, separate tests between the two motivating conditions were carried out at the High and Low levels. The mean Test 3 minus Test 2 difference of .102 for the IM-High subgroup differs significantly from the mean difference of −.075 for the Control-High subgroups at the .5% level ($F = 10.51$, 1 and 92 *df*) while the mean differences of .044 and .084 for the IM-Low and Control-Low subgroups, respectively, yields an F less than 1. Of the Within *Ss* components of variance only the L × S and M × L × S × R interactions reach significance at the 5% level.

The values of the theoretical parameters, $Pr(D_L \geq d)$, $Pr(D_L \leq -d)$, $Pr(D_N \geq d)$, and $Pr(D_N \leq -d)$, obtained from Equations 3–10 in conjunction with the data of Table 2, are shown in Table 5. In solving

TABLE 5

DERIVED VALUES OF $Pr(D_L \geq d)$, $Pr(D_L \leq -d)$, $Pr(D_N \geq d)$ AND $Pr(D_N \leq -d)$ FOR THE TWO LEVELS OF THE INCREASED MOTIVATION AND CONTROL GROUPS ON TESTS 2 AND 3, CLASSIFIED IN TERMS OF STIMULI AND DIRECTION OF DIFFERENCE

Level of Motiv.	Stimuli	Test	Increased Motivation		Control	
			Direction of Difference		Direction of Difference	
			$(D \geq d)$	$(D \leq -d)$	$(D \geq d)$	$(D \leq -d)$
High	Letters	2	.333	.017	.511	.033
		3	.434	.050	.391	.055
	Numbers	2	.466	.032	.533	.051
		3	.628	.112	.527	.009
Low	Letters	2	.277	.135	.280	.094
		3	.176	.122	.194	.040
	Numbers	2	.318	.218	.323	.163
		3	.268	.160	.257	.135

TABLE 6

PREDICTED PROPORTION OF CORRECT AND INCORRECT RESPONSES BY THE TWO LEVELS OF THE INCREASED MOTIVATION AND CONTROL GROUPS TO THE COMPOUND STIMULI FOR TESTS 2 AND 3

	Increased Motivation				Control			
	High		Low		High		Low	
	Test 2	Test 3	Test 2	Test 3	Test 2	Test 3	Test 2	Test 3
$Pr(LN, +)$	.798	.816	.592	.568	.845	.825	.635	.615
$Pr(LN, +\|L, +; N, +)$ $[Pr(L, +)\cdot Pr(N, +)]$	.431	.508	.280	.242	.524	.475	.304	.265
$Pr(LN, -\|L, +; N, +)$ $[Pr(L, +)\cdot Pr(N, +)]$	.041	.017	.034	.050	.024	.032	.040	.058
$Pr(LN, +\|L, +; N, -)$ $[Pr(L, +)\cdot Pr(N, -)]$	.130	.097	.129	.115	.143	.125	.135	.130
$Pr(LN, -\|L, +; N, -)$ $[Pr(L, +)\cdot Pr(N, -)]$	.057	.070	.128	.120	.048	.036	.114	.123
$Pr(LN, +\|L, -; N, +)$ $[Pr(L, -)\cdot Pr(N, +)]$	.196	.194	.149	.161	.154	.193	.156	.162
$Pr(LN, -\|L, -; N, +)$ $[Pr(L, -)\cdot Pr(N, +)]$	.049	.039	.087	.101	.039	.059	.080	.076
$Pr(LN, +\|L, -; N, -)$ $[Pr(L, -)\cdot Pr(N, -)]$	.041	.017	.034	.050	.024	.032	.040	.058
$Pr(LN, -\|L, -; N, -)$ $[Pr(L, -)\cdot Pr(N, -)]$	.056	.058	.159	.161	.044	.048	.131	.127
Σ (Obs.-Pred.)	.004	−.060	−.055	.007	−.018	−.039	−.058	.030
$\sqrt{\frac{\Sigma (\text{Obs.-Pred.})^2}{4}}$ for (LN, +)	.012	.022	.037	.015	.010	.010	.022	.016
(LN, −)	.005	.017	.037	.017	.012	.010	.021	0.16

for the values of Table 5, all eight equations were used in contrast to the method previously employed (**4**). The values of Table 5 were substituted in Equations 3–10 to arrive at the predicted values of Table 6. The adequacy of the fit is indicated by the difference in observed and predicted scores and the root mean squares of the differences presented at the bottom of Table 6.

With some confidence in the fit of the model to the data, an examination of the shifts in the theoretical values of Table 5 is in order. For the IM-High subgroup all four values increase from Test 2 to Test 3. However, with the exception of the shift in $Pr(D_L \leq -d)$ for the Control-High subgroup all other values of Table 5 decrease from Test 2 to Test 3. An analysis of variance, summarized in Table 7, was carried out on the factorial combination of the Motivation Conditions, Test 2 Performance Levels, Stimuli (Letters and Numbers), and Response Direction ($D \geq d$ and $D \leq -d$) dichotomies to evaluate the Test 2 to Test 3 shifts. This analysis indicates no significant differences attributable to the Stimuli or Response Direction factors, either as main effects or in interactions. Accordingly, the Within Subgroups variation was pooled to provide the error term for testing the Between Subgroups components, each of which is significant beyond the 5% level.

TABLE 7

SUMMARY OF THE ANALYSIS OF VARIANCE ON THE TEST 3 MINUS TEST 2 DIFFERENCES FOR DERIVED VALUES OF $Pr(D_L \geq d)$, $Pr(D_L \leq -d)$, $Pr(D_N \geq d)$ AND $Pr(D_N \leq -d)$

Source	*df*	*F*
Between subgroups	3	
Motivation conditions (MC)	1	8.35*
Levels (L)	1	13.78**
MC × L	1	7.61*
Within subgroups	12	(.002135)
Stimuli (S)	1	0.62
Direction of difference (DD)	1	0.16
MC × S	1	0.005
MC × DD	1	0.86
L × S	1	0.16
L × DD	1	0.53
S × DD	1	1.12
MC × L × S	1	0.14
MC × L × DD	1	1.00
MC × S × DD	1	0.05
L × S × DD	1	0.15
MC × L × S × DD	1	(.004424)

Note.—Mean squares for error in parentheses.
* $P = .05$.
** $P = .01$.

The over-all mean difference in shift for the IM group is .019 and for the Control group, −.048. Separate tests between the two motivating conditions show that the IM-High mean shift of .094 differs from the Control-High mean shift of −.036 at the .5% level of significance ($F = 15.95$, 1 and 12 *df*), whereas for the Lows, the IM and Control mean shifts of −.056 and −.058, respectively, yield an F less than 1. This pattern of significance of differences between the motivating conditions for the two levels of performance on Test 2 is the same as was found in the analysis of the Table 3 data.

DISCUSSION

The present study, designed to examine the effects of a shift in motivation on a complex task, points up the intricacies of such investigations. By definition, a complex task is one in which more than one response tendency may be operative in producing the response and, as such, calls for an analysis of the response hierarchy if successful predictions are to be made. Such an analysis was made in the present study by a model for response tendency combination, and the fit of the model to the data appears quite satisfactory.

The model, which, in conjunction with the Hullian postulate of $D \times H$,

forms the basis for the analysis of the data, suggests that the direction of the gross shifts of Pr(L,+), Pr(N,+) and Pr(LN,+) as a function of change in D are indeterminate because of compensating effects among the components. That no uniform direction of shift occurred for these values in the present data is shown clearly in Table 2.

The model does indicate, however, that consistent, differential shifts for the IM and Control groups from Test 2 to Test 3 should be found in two aspects of the data. Changes in D, presumed to be effected by the differential instructions to the two groups, should be manifested by differential shifts for the groups in the theoretical values, $Pr(D_L \geq d)$, $Pr(D_L \leq -d)$, $Pr(D_N \geq d)$, and $Pr(D_N \leq -d)$, and in certain categories of the compound stimuli. Analysis of the data tend to bear out the theory in both instances.

An over-all increase in the theoretical values for the IM group as a function of increased D is as expected from the theory while the over-all decrease in these values for the Control group suggests that the instructions to this group may have been D decreasing. The data from the compound stimuli relevant to the hypothesis,

$$Pr(LN,+|L,+;N,+) \quad [Pr(L,+)\cdot Pr(N,+)],$$
$$Pr(LN,-|L,+;N,+) \quad [Pr(L,+)\cdot Pr(N,+)],$$
$$Pr(LN,+|L,-;N,-) \quad [Pr(L,-)\cdot Pr(N,-)]$$

and

$$Pr(LN,-|L,-;N,-) \quad [Pr(L,-)\cdot Pr(N,-)]$$

show a pattern of shift for the IM group predicted by an increased D while the Control group tends to show no shift.

Analyses on both sets of data, however, show an interaction effect between the motivating conditions and the level of performance on Test 2. It is clear that the theory is supported in both cases by the results from the IM-High and Control-High subgroups. The results for the IM-Low and Control-Low subgroups are not equally clear since there appears to be no difference between these two subgroups in either analysis. This result would not be at variance with the theory, however, if it could be shown either that the motivating instructions affect the High and Low subgroups differentially or that D_L and D_N for the Low subgroups are not sufficiently large to be affected significantly by the change in D employed.

Summary

An hypothesis, relating to the effects of a shift in motivation on the performance of a complex learning task, was derived from the Hullian postulate of $D \times H$ in conjunction with a model for response tendency combination. A total of 96 men were divided into two groups and allowed equal opportunity to learn to categorize letter pairs and number pairs into one of two response classes. Both groups were tested on the letter and number pairs and on letter-number compound stimuli before and after differential motivating instructions. Analyses of the shifts in performance for the two groups on two aspects of the data—the theoretical parameters of the model and the relevant categories of the compound stimuli—support the use of the model in gaining information about the response hierarchy and also the general hypothesis of a direct relationship between D and specific response tendencies. The results emphasize the importance of a knowledge of the response hierarchy in the experiments dealing with complex learning tasks since gross measures of performance may cover up or disguise systematic variation in the components of these measures.

References

1. Alper, T. C. Task-orientation vs. ego-orientation in learning and retention. *Amer. J. Psychol.*, 1946, **59**, 236–248.
2. Atkinson, J. W. Towards experimental analysis of human motivation in terms of motive, expectancy, and incentive. In J. W. Atkinson (Ed.), *Motives in Fantasy, Action and Society*. Princeton: Van Nostrand, 1958.
3. Axelrod, H. S., Camen, E. L., & Heilizer, F. The correlates of manifest anxiety in stylus maze learning. *J. exp. Psychol.*, 1956, **51**, 131–138.
4. Birch, D. A model for response tendency combination. *Psychometrika*, 1957, **22**, 373–380.

5. Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.
6. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
7. Hilgard, E. R., Jones, L. V., & Kaplan, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.
8. Lazarus, R. S., Deese, J., & Hamilton, R. Anxiety and stress in learning: The role of intraserial duplication. *J. exp. Psychol.*, 1954, **47**, 111–114.
9. Lazarus, R. S., & Ericksen, C. W. Effects of failure stress upon skilled performance. *J. exp. Psychol.*, 1952, **43**, 100–105.
10. Lucas, J. The interactive effects of anxiety, failure, and intraserial duplication. *Amer. J. Psychol.*, 1952, **65**, 166–173.
11. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
12. Ramond, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.
13. Russell, W. A. Retention of verbal material as a function of motivating instructions and experimentally induced failure. *J. exp. Psychol.*, 1952, **43**, 207–216.
14. Sarason, I. Effect of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253–260.
15. Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.
16. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
17. Spence, K. W., & Farber, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, **47**, 127–134.
18. Spence, K. W., Farber, I. E., & McFann, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.
19. Spence, K. W., Taylor, J., & Ketchel, R. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.
20. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
21. Taylor, J. A. A personality test for manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(February 10, 1958)

Journal of Experimental Psychology
Vol. 56, No. 6, 1958

THE ACQUISITION OF IMITATIVE AND OPPOSITION RESPONSES UNDER TWO CONDITIONS OF INSTRUCTION-INDUCED SET[1]

VERA T. KANAREFF AND JOHN T. LANZETTA

Fels Group Dynamics Center, University of Delaware

In several recent studies of imitative behavior (**6, 7**) a lower proportion of imitative responses during acquisition was obtained for adult *Ss* than in the early Miller and Dollard studies of children and rats (**5**). To account for this poorer performance, Schein (**6**) proposed that adults may associate guilt or anxiety with imitative behavior in a "test" situation. A series of studies by Luchins (**4**) indirectly supports Schein's speculation by demonstrating that acquisition of imitation in older children may also be affected by guilt or expectations of punishment. In Luchins' studies, even when a confederate gave obviously true answers, not all *Ss* consistently agreed with these answers. Confirmation of the partner's responses and clarification of the evidence were attempted in order to obtain 100% agreement, but consistent agreement resulted only when *Ss* were given some means for direct measurement. Subsequent interviews indicated that those *Ss* who did not imitate consistently thought they might "get into trouble" or be considered "cheats" if they did imitate.

The above results suggest at least two variables that may affect the rate and level of acquisition of an imitative response in a choice situation: the instrumental values of the imitative response and the other response alternatives and the anticipated social consequences (e.g., social approval or disapproval) associated with these behavior alternatives. Conflict will presumably arise in those situations in which a given behavior pattern such as imitation is instrumental to goal achievement but where negative sanctions for that behavior pattern are anticipated. For example, either expectations that imitation is not legitimate or anxiety previously conditioned to imitation may be elicited in test-like situations, and yet in some of these test-like situations imitation may be the only effective means for doing well on the task. Since expectations of disapproval should increase as the number of imitative responses increases one might anticipate that a general method of resolution of such conflicts would be an oscillating behavior sequence. The total number of imitative responses in these sequences would be less than would be obtained under neutral or positive sanctions but greater than would be obtained if imitation were no more instrumental to goal achievement than any other response alternative.

In the present investigation, the acquisition of imitation and opposition responses in adult *Ss* is examined under two instruction-induced sets differing in the extent to which guilt for imitative responses would be elicited: one set of instructions was designed to induce negative sanctions

[1] The writers wish to express their appreciation to Stuart Cook for his discussion concerning the variables, to Jo Anne Davis for her assistance in collecting and analyzing the data, and to Seymour Berger for his assistance in conducting pilot research on this problem.

towards imitative behavior, the other was designed to induce neutral or slightly positive sanctions. In addition, the instrumental value of imitation was manipulated by varying the percentage of partner's choices which were considered correct; i.e., the three probabilities of occurrence of reinforcement for imitation were .80, .50, and .20.

The primary hypothesis under investigation may be stated as follows: As the instrumental value of imitation increases, the rate and level of acquisition of an imitative response will increase; the extent of this increase, however, will be a decreasing function of the strength of the prevailing negative sanctions towards imitation. Thus, the extent of increase in the rate and level of acquisition of imitation as the probability of reinforcement for imitation increases from .20 to .80 should be greater under the instructions which induce neutral or slightly positive sanctions towards imitation than under the instructions which induce negative sanctions towards imitation.

In addition to a test of this major hypothesis several subsidiary hypotheses concerning the resolution of the induced conflict will be explored. Under conditions whereby imitation is negatively sanctioned but is instrumental to goal achievement, a conflict will be engendered; *Ss* must choose between conforming to social norms and performing poorly on the task or employing a nonsanctioned behavior pattern and performing well. Those *Ss* who imitate may reduce the conflict or the anxiety associated with it by denying the existence of the negative sanctions, underestimating the extent of imitation, or underestimating their own success; whereas those *Ss* who do not imitate may reduce the conflict or anxiety by lowering their level of aspiration, overestimating their own success, underestimating their partner's success, denying the importance of the immediate task, or any combination of the above. Questionnaire data will provide exploratory information for the consideration of these subsidiary hypotheses.

Method

General Procedure

The *Ss*, in pairs, performed a pitch discrimination task ostensibly designed to improve existing methods for the identification of underwater craft. No communication between *Ss* was possible during the actual test period, for they were seated in separate, partially enclosed booths. A small metal response box containing two response keys and a pair of signal lights was located in each booth. The *S* indicated his judgment on each pitch discrimination by pressing the appropriate key on the response box. The two *Ss* were told they would give their judgments in a fixed order and that their partner's judgment would be indicated by the colored lights above his response keys. Since both *S*'s were instructed that their partner would give his judgment first and since the response box lights in both of the booths were connected to a central control console, *E* could feed in a programmed sequence of "partner's" judgments to each *S*. In post-session debriefings, *Ss* revealed no suspicions that the *E*-controlled responses of the partner were not those of the actual partner. The correctness of *S*'s choice was indicated by the presence or absence of a white light located on the front panel of his booth; this was controlled by *E* from the central console. In 60 pairs of tones (the critical trials) both the standard and the comparison tone were the same; there were 15 pairs of tones each at 800, 900, 1000, 1100 cps. In 15 other pairs of tones the second tone differed in pitch from the standard tone; the standard tone was either 800, 900, 1000, or 1100 cps while the second tone was either above or below the standard. The pairs of different tones were randomly interspersed among the critical trials to prevent *Ss* from suspecting that both tones in a pair might be the same. A random sequence of the entire set of 75 pairs of tones was recorded on tape.

The experimental procedure is readily apparent in that portion of the instructions which follows the introductory setting: "In this session two different tones will be presented, one at a

time, over the loudspeaker. Your task is to judge whether the second tone is higher or lower than the first one. The two tones will always be different, that is, the second tone is always either higher or lower than the first one. The differences, however, may be very small or imbedded in a noisy background, for we are particularly interested in situations where judgments are very difficult to make. It is in situations such as these that the number of judges will be critical. If there are times that you do not detect a difference, make a guess. The difference may be so subtle that you may not be aware that you noticed it. Although the discriminations are difficult, *most* of our subjects readily learn to make finer judgments with practice and get at least 80% of the judgments correct. Some people, however, can learn much more rapidly than others."

At this point *S*s were sent to separate, partially enclosed booths to read additional instructions. The instructions in the booths were identical for *both* *S*s; i.e., each *S* was told to give his judgments after his partner gave his, as follows (in this manner, *E* could control the sequence of "partner's" responses and run two *S*s simultaneously):

"The person sitting in this booth is the second person in your group—Subject B. Briefly, this is the procedure you are to follow: (*a*) First you will hear three low-pitched tones in rapid succession. This will be warning signal 1. (*b*) A few seconds after warning signal 1, the two tones to be judged will be presented, one after the other. (*c*) *Your partner*, Subject A, will make his judgment immediately. His judgment must be made before warning signal 2 is presented. Warning signal 2 consists of three high-pitched tones in rapid succession. (*d*) You will have ten seconds *after* warning signal 2 to make *your* judgment. (Please do *not* make your judgment *before* warning signal 2. The experimenter will be unable to record your judgment unless it is made *after* warning signal 2.) (*e*) Ten seconds after warning signal 2 you will again hear warning signal 1 to alert you for the next judgment.

"Use the small metal box located in your booth to indicate your judgment. Press the button at the bottom left (marked 'higher') if you judge the second tone to be higher than the first; press the button on the bottom right (marked 'lower') if you judge the second tone to be lower than the first.

"The row of lights above these buttons represents your partner's choices. When the green light (located above the 'higher' response button) goes on, it means that your partner, Subject A, has judged the second tone to be higher than the first. When the red light (located above the 'lower' response button) goes on, it means that he has judged the second tone to be lower than the first.

"After you have made your judgment, *watch* the light on the panel wall in front of you. If your judgment was correct, a white light will flash on; if your judgment was incorrect, the white light will *not* appear."

The *E* then briefly summarized the task procedure, gave *S*s five practice trials without reinforcement, and then proceeded with the test series.

Experimental Conditions

Percentage of "partner's" judgments called correct (probability of reinforcement).—The sequence of judgments considered correct for the 60 pairs of identical tones and the corresponding three response sequences for the "partner" were randomized within blocks of 10 trials with the restriction that there be an equal number of higher and lower judgments. Where equal distribution of responses among the higher and lower judgments was not possible, the odd judgments were balanced over blocks of trials. The three "partner"-response sequences form the three probabilities of reinforcement, .80, .50, and .20; i.e., in one sequence, 80% of the "partner's" judgments are correct; in another sequence, 50% of them are correct; in the third sequence, only 20% of them are correct. These probabilities of reinforcement are taken with respect to the 60 critical trials. Since the "partner's" judgments were always correct on the remaining 15 trials, the probabilities of reinforcement for the entire sequence are .84, .60, and .36. Whether *S* was correct or not depended on whether he agreed with the "partner" or disagreed with him and whether the "partner's" judgment happened to be correct or incorrect on that particular trial. If *S* agreed consistently with the "partner" throughout all the trials, he would have been correct 8 times in each block of 10 trials when 80% of the "partner's" judgments were correct, 5 times per block of 10 trials when 50% of the "partner's" judgments were correct, and 2 times per block of 10 trials when 20% of the "partner's" judgments were correct. Accordingly, imitation, or consistent agreement with the "partner," would be a behavior pattern instrumental to goal achievement in the .80 condition, and opposition, or consistent disagreement with the "partner," would be the instrumental behavior pattern in the .20 condition.

For the 15 pairs of different tones, the "partner" always judged correctly and *S* was correct if he agreed with his "partner."

Instruction-induced set.—That portion of the introductory setting common to both sets of

instructions was as follows: "Here is an illustration of the type of problem that we are studying. Our national and civil defense agencies have long been faced with the problem of identification of hostile and friendly underwater craft. Errors in judgment in identifying such underwater craft have led to instances where we have sunk our own submarines or allowed enemy craft to pass through our territory. Even our modern sonar equipment cannot correct for similar kinds of errors of judgment."

Negative sanctions towards imitative behavior were introduced in Instructions II by supplementing this common portion with the following: "We are attempting to determine whether the judgments made by a group of people are more reliable, more accurate, than the judgments of one person alone. We are, in addition, interested in the problem of how many people are needed to obtain accurate judgments. The problem may be posed as follows: Suppose that two people individually judge the same thing. How much difference in accuracy is there in their separate judgments and are their separate judgments more accurate than their combined judgments?"

In addition, the statement "This will be a test of your ability to discriminate tones in that we will see how well you do in comparison with each other" was included near the end of Instructions II, just prior to the practice series.

Neutral or slightly positive sanctions towards imitative behavior were introduced in Instructions I by supplementing the common portion with the following: "We are attempting to determine how much of an actual difference is necessary under varying conditions of distraction in order for people to recognize these differences. A related problem is the development of effective means for improving judgments that are difficult to make. One possibility is that working in a group may improve one's ability to distinguish differences when conditions for making such judgments are poor."

Experimental Design

The three probabilities of reinforcement and the two sets of instructions form a 2 × 3 orthogonal design. Twenty-four female and 24 male volunteers from the freshman English classes at the University of Delaware were randomly assigned in like-sexed pairs to the six cells of the design with the restriction that each cell contain two female pairs and two male pairs.

The number of imitative responses (judgments indicating agreement with the partner) per block of 10 trials in the critical series is the major variable under observation. A questionnaire assessing *S*'s aspiration level and estimates of his success, of his partner's success, and of the amount of agreement with the partner was also used to provide exploratory data about *S*'s expectations in the experimental situation. A continuous rating scale ranging from 100% to 0%, or "applies very much" to "doesn't apply at all," accompanied the following questions and statements in the questionnaire:[2]

1. Before you even began to make any judgments, what percentage did you expect to judge correctly?
2. During the period of actually making judgments, what percentage did you expect to judge correctly?
3. To what extent were you certain that you'd reach the level checked in question 2?
4. On each trial, I made a very strong effort to keep from making a mistake.
5. On the whole, I made a very strong effort to keep from getting below the level checked in question 2.
6. What percentage of your judgments did you judge correctly?
7. What percentage of his judgments did your partner judge correctly?
8. What percentage of your judgments were in agreement with those of your partner?
9. To what extent did you feel frustrated during the session?
10. To what extent did you feel that you were competing with your partner?

Results

Imitative Responses

An analysis of variance and orthogonal polynomial analyses of trend (**3**) were applied to the number of imitative responses per block of ten trials in the series of 60 critical trials.

The summary of the analysis of variance is presented in Table 1. Results from the Duncan Range test (**1**) on the Instructions × Probability of Reinforcement interaction means are given in Table 2. The three probability of reinforcement means under Instructions I differ significantly from one another; the largest amount of imitative responses occurs when imitation is instrumental to

[2] Copies of this questionnaire are available from the authors upon request.

TABLE 1

ANALYSIS OF VARIANCE OF THE IMITATIVE RESPONSES

Source	*df*	*MS*	*F*	*P*
Instructions	1	1.00		
Probability Ratio	2	54.19	12.04	
Trials	5	.08		
Sex	1	.00		
I × PR	2	43.92	9.76	<.01
I × T	5	4.61	2.11	
I × S	1	.43		
PR × T	10	5.01	2.30	<.05
PR × S	2	2.97		
T × S	5	2.31	1.06	
I × PR × T	10	3.62	1.66	
I × PR × S	2	2.36		
PR × T × S	10	1.93		
I × T × S	5	1.56		
I × PR × T × S	10	2.84	1.30	
Ss within	36	4.50	2.11	<.01
Ss within × Trials	180	2.18		
Total	287			

Note:—The *Ss*-within cells is used as the error term for nonrepeated measures; *Ss* within × Trials is used as the error term for all others. The significant terms are then tested against the appropriate significant higher-order interactions where such interactions exist.

goal achievement (.80 reinforcement) and the least amount occurs when opposition is instrumental (.20 reinforcement). None of the probability of reinforcement means, however, differ under Instructions II. In addition, the two instructions means differ significantly within the .80 and .20 reinforcement conditions but not within the .50 reinforcement condition: There are more imitative responses under the instructions inducing neutral or slightly positive sanctions towards imitation (Instructions I) than there are under the instructions inducing negative sanctions towards imitation (Instructions II) when imitation is instrumental; there are less imitative responses under Instructions I than under Instructions II when opposition is instrumental; and there are no significant differences between instructions in the amount of imitative responses when neither imitation nor opposition is instrumental.

TABLE 2

DUNCAN RANGE TEST RESULTS ON THE INSTRUCTIONS BY PROBABILITY OF REINFORCEMENT IMITATION MEANS*

Instructions	Probability of Reinforcement		
	.20	.50	.80
I	4.46[a]	5.44[b]	7.27[a]
II	5.54[b]	5.58[b]	5.69[b]

* Those means marked with a letter in common do not differ significantly at the .05 level.

The Duncan Range test results on the Probability of Reinforcement × Trials interaction means in Fig. 1 indicate that the three curves differ from each other, since on at least one block of trials the means differ significantly.

Orthogonal analyses of trends were applied to the data from Instructions I, the data from Instructions II, and also the combined instructions data. Only the summaries of the analyses of the Instructions I data and the Instructions II data are presented in Tables 3 and 4. As expected, the Between Group Means, the Probability of Reinforcement, and the

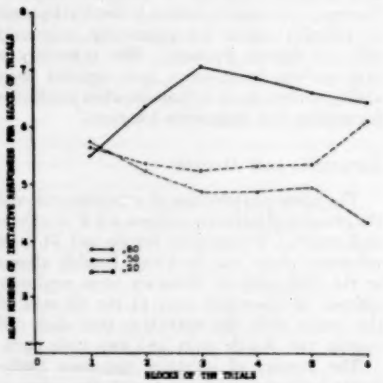


FIG. 1. Mean number of imitative responses per block of trials as a function of probability of reinforcement.

Probability of Reinforcement × Instructions terms are highly significant in the orthogonal polynomial analysis of trends in the combined data. The Between Group Trends is significant as well as the linear and quadratic components of these trends. There are also significant differences in trend among the different probabilities of reinforcement in both the linear and the quadratic components. The differences in the trends among the probabilities of reinforcement are evident in Fig. 1: When imitation is instrumental, the number of imitative responses at first increases over trials and then drops slightly in contrast to the immediate drop in imitative responses when opposition is instrumental.

In addition, the quadratic component of the differences in trend among the Instructions × Probability of Reinforcement groups is significant. These differences in trends are evident in Fig. 2. The differences among the probability of reinforcement curves under Instructions I are striking. When imitation is instrumental, the number of imitative responses increases with trials and then levels off; when opposition is instrumental, the number of imitative responses decreases with trials and then levels off; when neither imitation nor opposition is instrumental, slightly more than half the responses on each block of trials are imitative responses. Under Instructions II, both the .80 and .20 reinforcement conditions show

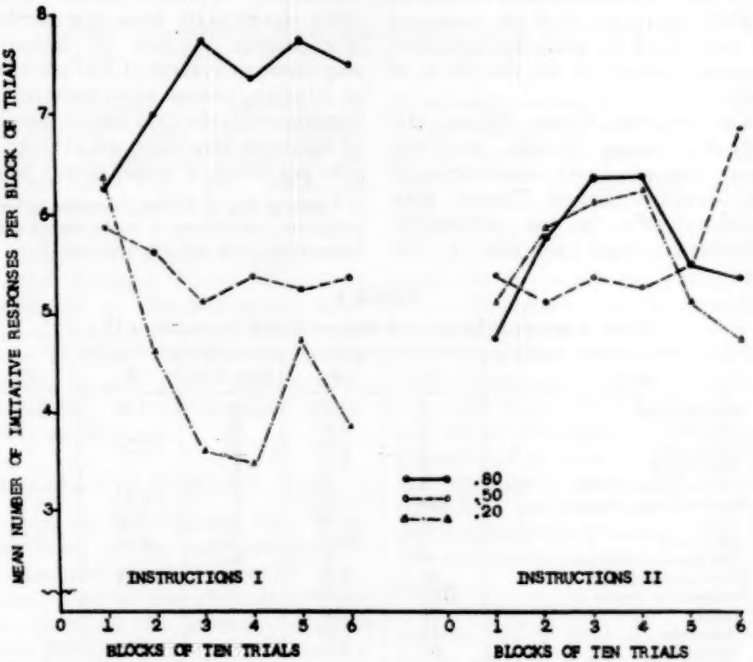


FIG. 2. Mean number of imitative responses per block of trials as a function of probability of reinforcement and instructions.

TABLE 3

TREND ANALYSIS OF IMITATIVE RESPONSES UNDER INSTRUCTIONS I

Source	*df*	*MS*	Error Term	*F*	*P*
A. Over-all trend	(5)	(1.91)	E	1.16	
1. Linear	1	2.14	E_1	1.40	
2. Quadratic	1	3.33	E_2	1.60	
3. Deviation from 1 and 2	3	1.36	E_3		
B. Between group means	2	97.84	D	23.86	<.01
C. Between group trends	(10)	(5.24)	E	3.18	<.01
1. Linear	2	11.93	E_1	7.80	<.01
2. Quadratic	2	9.41	E_2	4.52	<.05
3. Deviations from 1 and 2	6	1.62	E_3	1.05	
D. Between ind. means	21	4.10	E	2.48	<.01
E. Between ind. trends	(105)	(1.65)			
1. Linear	21	1.53			
2. Quadratic	21	2.08			
3. Deviations from 1 and 2	63	1.54			
Total	143				

an early increase and subsequent decrease in imitative responses, whereas in the .50 reinforcement condition, slightly more than half the responses on each block of trials are imitative responses except on the last block of trials.

The Between Group Means, the Between Group Trends and the linear and quadratic components of the Between Group Trends give significant *F*'s in the orthogonal polynomial trend analysis of the Instructions I data.[3] Duncan Range test results indicate that the three probability of reinforcement means differ significantly from one another. The largest amount of imitative responses, an average of 7.27 per block of 10 trials, occurs when imitation is instrumental; the next largest amount of imitative responses, an average of 5.44 per block of trials, occurs when

[3] Refer to Fig. 2 for the differences in trend among the probabilities of reinforcement under Instructions I and under Instructions II.

TABLE 4

TREND ANALYSIS OF IMITATIVE RESPONSES UNDER INSTRUCTIONS II

Source	*df*	*MS*	Error Term	*F*	*P*
A. Over-all trend	(5)	(2.78)	E	1.01	
1. Linear	1	1.80	E_1		
2. Quadratic	1	6.79	E_2	2.34	
3. Deviations	3	1.77	E_3		
B. Between group means	2	.27	D		
C. Between group trends	(10)	(3.39)	E	1.24	
1. Linear	2	4.47	E_1	1.56	
2. Quadratic	2	12.08	E_2	4.17	<.05
3. Deviation	6	.13	E_3		
D. Between ind. means	21	4.13	E	1.51	<.05
E. Between ind. trends	(105)	(2.74)			
1. Linear	21	2.87			
2. Quadratic	21	2.90			
3. Deviations	63	2.64			
Total	143				

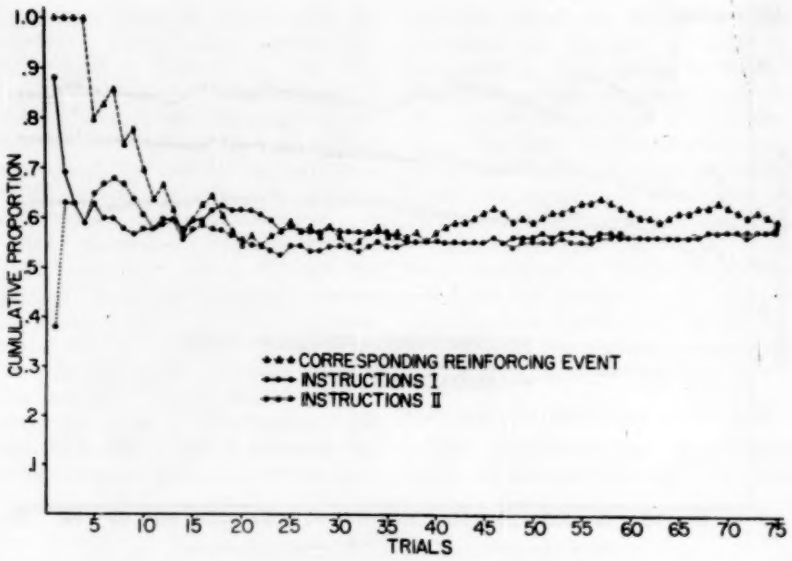


FIG. 3. Cumulative proportions of the imitation response and its reinforcing event in the .60 probability of reinforcement condition.

neither imitation nor opposition is instrumental; and the smallest amount of imitative responses, an average of 4.46 per block of trials, occurs when opposition is instrumental.

Although the quadratic components of the Between Group Trends is significant in the orthogonal polynomial trend analysis of the Instructions II data, the Between Group Trends is not.[4] Nor do the three probability of reinforcement means differ significantly.

Event-Matching Behavior

Estes (2) has suggested that learning theory models could also be used to describe learning in social situations. One of the deductions from his model is the general matching law: The cumulative proportions of a given response and the corresponding reinforcing event tend to equality as the number of trials approaches infinity. The present study suggests that some modifications in the model may be necessary as Ss' response probabilities in certain of the conditions did not reach the probability of reinforcement for the response, but stabilized some distance below it. The cumulative proportions of the imitative response and the corresponding reinforcing event, E_1, for the .50 and .80 reinforcement conditions are plotted in Fig. 3 and 4, respectively; the cumulative proportion of the opposition response and the corresponding reinforcing event, E_2, for the .20 reinforcement condition are plotted in Fig. 5.

Whenever the "partner's" choice is correct, E_1 is said to occur, and E_2 is said to occur whenever the "partner's" choice is incorrect; thus,

[4] See Footnote 3.

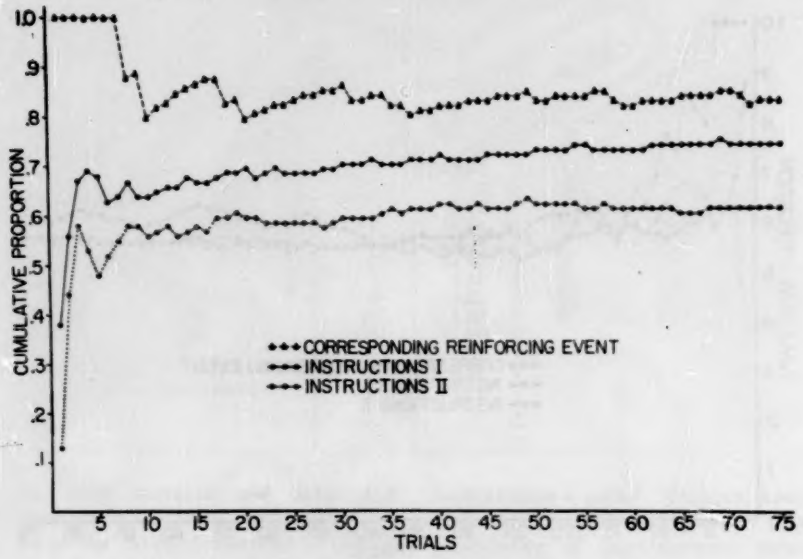


FIG. 4. Cumulative proportions of the imitation response and its reinforcing event in the .84 probability of reinforcement condition.

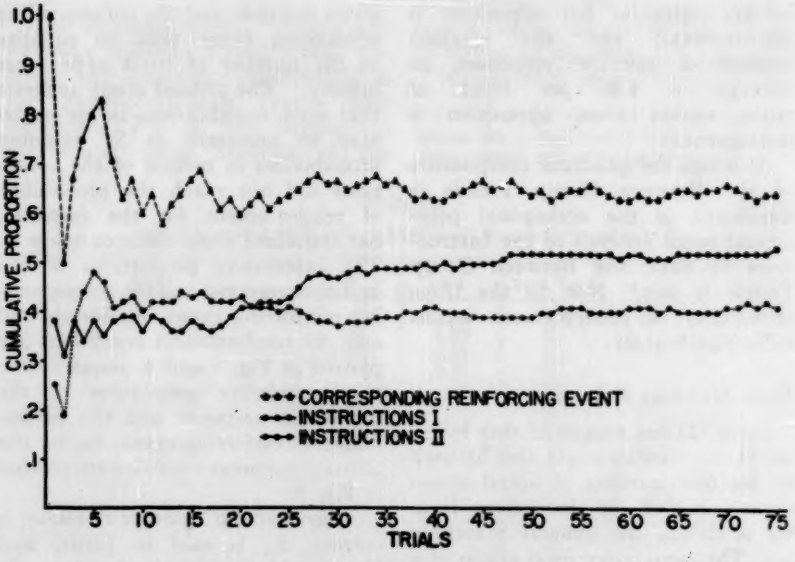


FIG. 5. Cumulative proportions of the opposition response and its reinforcing event in the .36 probability of reinforcement condition.

E_1 occurs when S agrees with his "partner" and is called correct (the white light is flashed on) and also when S opposes his "partner" and is called incorrect (the white light is not flashed on), whereas E_2 occurs when S agrees with his "partner" and is called incorrect, and also when he opposes his "partner" and is called correct.

The cumulative proportions are taken over the entire 75-trial sequence. Since the "partner's" choice is correct on all 15 of the noncritical trials, the probabilities of reinforcement for the entire sequence are .84, .60, and .36. Hence, the cumulative proportion of E_1 stabilizes around .60 and .84 in Fig. 3 and 4, respectively, while the cumulative proportion of E_2 stabilizes about .64 in Fig. 5. The matching law appears to hold for both sets of instructions in the .60 probability of reinforcement. For the .36 and .84 probabilities of reinforcement, however, the cumulative proportion of the given response doesn't reach the cumulative proportion of its corresponding event but stabilizes at about .54 and .74, respectively, when neutral or slightly positive sanctions for imitation are induced (Instructions I) and about .41 and .61, respectively, when negative sanctions are induced (Instructions II).

Questionnaire Responses

The intercorrelations among the items of the questionnaire are pre-

TABLE 5

INTERCORRELATIONS AMONG THE QUESTIONNAIRE ITEMS

Probability of Reinforcement	Questionnaire Items	2	3	4	5	6	7	8	9	10
.20	1	.49*	.54*	.23	.17	.57*	.23	.54*	−.22	.18
	2		.91**	.20	.31	.89**	.45	.17	−.53*	−.33
	3			.27	.37	.79**	.56*	.17	−.59*	−.23
	4				.52*	.18	.19	.04	−.03	.14
	5					.10	.16	.04	−.09	.22
	6						.48	.31	−.34	−.26
	7							.34	−.27	.03
	8								−.30	.10
	9									.40
.50	1	−.03	−.17	−.23	−.27	.17	.17	.41	.19	−.08
	2		.44	.42	.29	.06	.43	.38	.40	.39
	3			.28	.34	.53*	.46	.11	−.06	.02
	4				.09	.01	.25	.34	.49	.02
	5					.12	.28	.18	.53*	.10
	6						.37	.25	−.02	.15
	7							.16	.32	.43
	8								.31	.04
	9									.08
.80	1	.62*	.25	.40	.33	.72**	.65**	.68**	−.59*	.37
	2		.26	.64**	.42	.64**	.72**	.58*	−.62*	−.42
	3			.21	.20	.34	.43	.52*	−.33	−.27
	4				.52*	.35	.15	.20	−.57*	−.33
	5					.31	.34	.22	−.06	−.18
	6						.72**	.82**	−.75**	−.57*
	7							.83**	−.46	.41
	8								−.57*	−.43
	9									.34

* $P = .05$.
** $P = .01$.

sented by probability of reinforcement in Table 5; the pattern of significant relationships among questionnaire items clearly varies with probability of reinforcement. When neither imitation nor opposition is instrumental, the only significant correlations are between confidence in attaining the level of aspiration and estimate of own success and between over-all effort exerted and reported frustration. When imitation is instrumental, however, a prominent cluster of significant intercorrelations exists among both levels of aspiration, the estimates of own success, of partner's success, and of agreement with the partner and low frustration. When opposition is instrumental, the prominent cluster includes both levels of aspiration, confidence in attaining the level of aspiration, and estimate of own success.[5]

[5] Since the significant results of the numerous additional analyses do not form an obvious pattern and are not easily interpretable, they will be mentioned only briefly. In the .20 condition, *Ss* underestimate percentage of agreement with a partner (19.36 mean deviation of estimated percentage of agreement from actual percentage of agreement) significantly more than *Ss* in the .50 condition (3.25 mean deviation), but not significantly more than those in the .80 condition (10.50 mean deviation); the mean deviations for the .50 and .80 conditions do not differ significantly. Imitation and overestimation of partner's performance are positively related ($r = .51$) in the .80 condition, whereas imitation and ongoing level of aspiration are positively related ($r = .37$) in the combined .80 and .20 conditions. Confidence in attaining one's level of aspiration is positively related to overestimation of one's own success ($r = .79$) in the .20 condition and negatively related with underestimation of agreement with the partner ($r = .-64$) in the .80 condition. In both the .80 and .20 conditions the ongoing level of aspiration and overestimation of own success are positively related ($r = .59$), and confidence in attaining the level of aspiration is positively related with overestimation of partner's success ($r = .53$).

Discussion

It was hypothesized that the utilization of a behavior most effective for reaching a task goal would depend upon the prevailing social sanctions. As predicted, expectations of approval or disapproval for imitation did affect the acquisition of imitative responses. When expectations of neutral or slightly positive social sanctions for imitation were induced, the rate and level of acquisition of imitative responses varied significantly with the instrumental value of the imitative response. When expectations of negative social sanctions were induced, these differences in rate and level of acquisition were depressed.

It is apparent from the data on the opposition responses, however, that the induced sets aroused expectations of social sanctions, not only for the utilization of imitative responses per se, but also for any utilization of the available information about another's behavior. In effect, the inductions modified *Ss*' set towards independence. In the condition where the utilization of information about others was perceived as legitimate (Instructions I) the two instrumental behaviors, imitation and opposition, were also socially sanctioned behaviors. In the condition where independence was perceived as the social norm (Instructions II) the social sanctions and instrumental values were conflicting; *Ss* were forced to choose between conforming to the social norms and performing poorly on the task or of employing a nonsanctioned behavior and performing well. These *Ss* apparently were aware of the elements forming the conflict situation. They were able to estimate the partner's performance as well as *Ss* in the condition where social norms and instrumental values were not conflicting,[6] and yet they did not imitate or oppose as consistently. It is possible that *Ss*

[6] Instructions did not yield a significant *F* ratio in the analyses of variance of estimates of partner's performance and in the deviation of partner's actual success from *S*'s estimate of the partner's success.

oscillated in their use of imitation or opposition when they were faced with the conflicting sanctions and instrumental values. From the methods of analysis employed, however, it was not possible to distinguish between an alternation behavior and completely independent behavior.[7]

It was also suggested that those *S*s who imitate may reduce the conflict, or the anxiety associated with it, by denying the existence of the negative sanctions, underestimating the extent of imitation, or underestimating their success, whereas those *S*s who do not imitate may reduce the conflict by either lowering their levels of aspiration, or overestimating their success, or underestimating their partner's success, or denying the importance of the immediate task, or any combination of the above. Only when imitation was instrumental did one of the predictions hold; imitation and overestimation of partner's performance were positively related. The questionnaire data, on the whole, did not substantiate these hypotheses.

Further analysis of the data was undertaken to examine the application of Estes' learning theory model to learning in social situations. Strict event-matching occurred only when neither imitation nor opposition was instrumental to success on the task, whereas, the event was undermatched when imitation or opposition was instrumental. One possible explanation is that there exists a generalized resistance to violating the notion that one can do well on his own without accepting outside assistance. The conflict between accepting possible failure by performing on one's own and being assured of success by copying or opposing another may then be resolved in such a way that *S*s as a group undermatch the event, the extent of the undermatching being determined by the prevailing social sanctions.

[7] An analysis of the number of runs of imitative responses indicates that the number of runs of imitative responses increases with increasing instrumental value of imitation, regardless of social sanctions.

Further research is needed to determine the conditions which affect the development of sanctions favorable to an imitative or independence set. Some of the variables to be considered are the characteristics of the partner, of the reward for goal achievement, and of the task. Studies of the effect of increasing the reward on the acquisition of imitative and opposition responses and the effect of varying the instrumental value on both the acquisition and extinction of imitative and opposition responses are in progress.

Summary

The present study tests the hypothesis that, although a given behavior—i.e., imitation or opposition—may be the most appropriate means for reaching an immediate task goal, the utilization of such behaviors depends on the prevailing social sanctions for its use.

Two *S*s, seated in separate booths, judged whether the second of a pair of identical tones in a series of 60 paired tones was higher than the first. Each *S* indicated his judgment by pressing the appropriate response key after his "partner's" choice was indicated by a light above the response keys; if *S*'s judgment was correct, a white light was flashed on. Since each *S* was instructed that his partner would judge first, *E* was able to present a standard randomized sequence of "partner's" choices, as well as a prearranged sequence of reinforcements for imitation responses. The two independent variables were: (*a*) Probability of reinforcement for imitation. The percentages of the "partner's" choices that were considered correct were either 80%, 50%, or 20%, i.e., the three probabilities of occurrence of reinforcement for imitative responses were .80, .50, and .20. (*b*) Instruction-induced set. One set of instructions was designed to induce expectations of neutral or slightly positive sanctions for imitation, while another set was designed to induce expectations of negative sanctions.

As predicted, the extent of increase in the rate and level of acquisition of imitation with increasing instrumental value was significantly greater under positive sanctions towards imitation than under negative sanctions. It was also apparent that the acquisition of opposition responses was affected by the instruction-induced set. This suggests that the instructions modified *S*'s set towards independence and not towards the utilization of imitative responses per se.

The questionnaire data did not substantiate

the subsidiary hypotheses on the modes of resolution of the conflict engendered in the condition where social sanctions and instrumental values were conflicting.

The hypothesis advanced by Estes that "event-matching" behavior occurs in social learning situations was substantiated only in the .50 probability of reinforcement condition. Undermatching of the event occurred in the .80 and .20 conditions, the extent of the undermatching being determined by the prevailing social sanctions.

References

1. Duncan, D. B. Multiple range and multiple F tests. *Biometrics*, 1955, **11**, 1–41.
2. Estes, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617.
3. Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.
4. Luchins, A. S. On agreement with another's judgment. *J. abnorm. soc. Psychol.*, 1944, **39**, 97–111.
5. Miller, N. E., & Dollard, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
6. Schein, E. H. The effect of reward on adult imitative behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 389–395.
7. Schwartz, N. Referred to in R. R. Bush and F. Mosteller, *Stochastic models for learning*. New York: Wiley, 1955. Ch. 12.

(Received for early publication
June 18, 1958)

Journal of Experimental Psychology
Vol. 56, No. 6, 1958

ALTERATIONS IN CRITICAL FLICKER FREQUENCY AS A FUNCTION OF AGE AND LIGHT:DARK RATIO[1]

ROSS A. McFARLAND, A. BERTRAND WARREN, AND CHARLES KARIS

Harvard School of Public Health and Northeastern University

Studies concerned with the behavioral assessment of the neurophysiological condition of the eye have employed, among other measures, indices of visual sensitivity. This index is frequently determined by employing threshold luminances for a target that may be exhibited (*a*) continuously, (*b*) for a brief duration as in a flash, or (*c*) intermittently as in flicker. Under well controlled test conditions, visual threshold determinations show a high degree of reproducibility and many investigators have considered that, especially among inexperienced *O*s, the response to flicker is more reliable than other types of visual tests (**9**).

In routine determinations of critical flicker frequencies, the *SD* from the mean is usually found to be less than 2 cps during an experimental session (**8, 13**). This stability of critical flicker frequency (CFF), under a standard condition of luminance, size, and retinal position of the test field, has led to the assumption that CFF may be regarded as a numerical entity and that deviations from the "normal" denote changes in the physiological state of the eye or the organism as a whole (**10, 11**).

One of the important organismic variables known to influence CFF variation among individuals is chronological age. Simonson, Enzer, and Blankstein (**14**), and Brozek and Keys (**2**) report a definite decrease in CFF for *O*s over 40, but little change in CFF was found before this age. Misiak (**12**) studied *O*s ranging in age from 7 to 89 yr. and found that average CFF declines rather regularly from age 20 to 89, but not without considerable individual variation. A more recent study by Coppinger (**4**) has established that the CFF-age function is linear and negative from the age of 20 to 90. Furthermore, the decreased sensitivity among older *O*s cannot be entirely attributed to diminished pupil size since Weekers and Roussel (**15**), using pupil dilation with atropine, find that there is still a CFF decrement with age.

Problem

While it has been generally accepted that CFF tends to be negatively correlated with age, there has been little uniformity in previous studies in the characteristics of the test field, especially in the light:dark ratio (LDR) of the flicker cycle. Furthermore, since only a few studies have used more than one LDR, little systematic information is available on the comparative sensitivity of the fusion threshold when the light-time proportion (LP) of the LDR is varied. Simonson, Enzer, and Blankstein (**14**) found an equal effectiveness of light-time proportions of 64%, 27%, and 17.5% among various age groups. Fisher (**6**) using light:dark ratios ranging from 2% to 98% reports that elderly *O*s (76 to 87 yr.) respond differently from

[1] This study is one of a series sponsored by the Commission on Accidental Trauma, Armed Forces Epidemiological Board, Department of Defense, and supported in part by funds from the Office of the Surgeon General, Department of the Army.

young *O*s (21 to 24 yr.) to the shortening of the dark interval of the flicker cycle.

The present experiment attempts to establish more precisely the nature of the relationship between foveal CFF and chronological age by (*a*) using a number of different light:dark ratios, and (*b*) employing a large sample of *O*s covering an age span of 76 yr.

METHOD

Observers.—Critical flicker frequency thresholds were obtained from 108 male *O*s ranging in age from 13 to 89 yr. Among them were secondary school and college students, office workers and clerks, taxi drivers, and residents of old-age homes. All were in good health and were nonindigent. The *O*s who ordinarily wore glasses were allowed to wear them during the test. The numbers in each age group were as follows:

Age Range	Number	Average Age
13–19	10	16.5
20–29	13	23.5
30–38	13	35.1
40–49	16	45.0
50–58	14	54.1
61–68	14	64.3
70–78	13	74.0
80–89	15	84.8

Apparatus.—In most studies on intermittent light stimulation, a motor-driven sector disc has been used for the production of flicker. With this type of apparatus the change of the LDR necessitated the use of different discs which interrupted the test light at a focal point in the optical system of the instrument. More recently, electronic devices have been used for flicker production. In comparison to the rotation devices, these have the advantages of being easier to operate and of eliminating noise and vibration, but they may have the disadvantage of having very short and fixed flash durations which do not change when frequency is altered (General Radio Strobotac).

For the present investigation, a flicker fusion apparatus[1] which activates a cold cathode light modulating tube (Sylvania R1131C) was used. The light pulses are of a square wave form, and the apparatus permits the adjustment of the duty cycle or operation time of the lamp to be varied from 2% to 98% of the repetition period. The flash frequency is variable from 1 to 125 flashes per second. The peak luminance is controlled by varying the amperage, and if necessary the tube can be illuminated continuously for use as a standard "steady" light.

The glow modulator tube was placed behind a piece of ground glass which provided a predominantly white test area of even illumination over its surface. The test area was circular with a diameter of 1 cm., and was in the center of a circular surround which had a diameter of 28 cm. The surround was illuminated by a concealed circular neon daylight white tube matched in hue to the test area. Two surround luminances were employed: 23.6 and .04 ml. Perforated shields and a Variac were used to change the surround luminance without a color change.

The test area and surround were mounted on the rear wall of a rectangular housing. The test area served as the fixation point and was 96 cm. in front of and level with *O*'s pupils. Thus, the test area subtended a visual angle of 36 min. of arc, while the surround subtended a visual angle of approximately 17°. These angular dimensions allowed the image of the test area to fall within the fovea. The surround extended 8.5° into the periphery.

During the threshold determinations, the position of *O* was comfortably but securely maintained by means of a modified dentist's chair with back, foot, shoulder, and arm rests. The positioning of *O* was accomplished by vertical and lateral chair adjustments by *E*. When *O*'s eyes were level with the test area, his head was then held in a fixed position by means of a chin and head rest.

Procedure.—Each *O* was first made familiar with the distinction between flicker and fusion by changing the frequency from steady to intermittent stimulation. When he understood the nature of the task, practice trials were given until the responses to change from "steady to flicker" and "flicker to steady" were consistent. To insure adaptation to the surround illumination, *O* was not allowed to look away from the test field during the preliminary trials. Binocular CFF's were then determined by the method of limits using discontinuous exposures. The duration of each exposure was constant at 1.5 sec., and the change of the flicker rate in ascending and descending determinations was constant at 1 cps.

In determining each individual threshold, readings were taken as the continuously in-

[1] Designed by W. Bruce Fisher while on sabbatical leave at the Harvard School of Public Health from Fresno State College and constructed to his specifications by the Grayson-Stadler Company, West Concord, Massachusetts. Model E6-22.

creased frequency led to the report of "steady," and also when the frequency was steadily decreased until *O* reported "flicker." During this process the luminance of the test area was kept constant while frequency was varied by *E*. As frequency was increased or decreased the light duration of the cycle varied proportionately. Consequently, the light-time fraction of the flicker cycle remained constant regardless of the frequency of intermittence. Threshold determinations for each *O* were obtained under 10 light-time proportions: 2%, 5%, 10%, 25%, 40%, 50%, 75%, 90%, 95%, and 98%.

In this study the peak current was kept constant for all threshold determinations. When the tube was illuminated steadily, this provided the test area with a mean luminance of 21.19 ml., as measured with a Macbeth illuminometer from the eye position of *O*.

For each LP level, two ascending and two descending thresholds were obtained. This provided CFF values in cycles per second for the frequency level at which flicker disappeared, and when flicker reappeared. The four determinations were averaged as the mean CFF.

During the first test session, which took approximately 1 hr., CFF's were determined with the 10 LDR's under two surround luminances: 23.6 and .04 ml. At these luminance levels, the high surround luminance is brighter than any of the uncompensated test area luminances, while the .04 ml. surround is dimmer than any of the test area luminances.

A second test session followed one week later in which the technique of testing was the same but the order of surround luminances was reversed. In both sessions the starting order for the various LP levels was random. For each *O*, 160 CFF's were determined. The values obtained from 108 *O*s, averaged for the two sessions, constitute the material presented in the following tables and figures.

Results

The means and *SD*'s of the fusion thresholds in cycles per second are presented in Tables 1 and 2. Inspection of the data for the various age groups shows that mean CFF

TABLE 1

Mean CFF's and *SD*'s for Eight Age Ranges at 10 Light: Dark Ratios When the Test Area Appears at the Center of a Surround of 23.6 ml.

Age Range		Light-Time Percentage in Flicker Cycle									
		2%	5%	10%	25%	40%	50%	75%	90%	95%	98%
13–19	*M*	44.82	46.17	47.65	47.25	45.12	44.52	38.06	29.84	20.72	10.45
	SD	1.77	1.83	1.84	2.27	1.92	1.80	2.57	2.21	1.75	2.13
20–29	*M*	42.83	44.39	46.00	46.00	44.37	42.43	37.84	29.85	20.88	10.67
	SD	2.02	1.94	2.19	2.56	2.63	2.77	2.09	2.44	3.11	2.39
30–38	*M*	41.12	42.81	44.27	44.83	42.79	42.84	37.71	29.49	20.19	9.88
	SD	2.02	1.92	2.27	2.72	2.60	2.75	2.64	3.18	2.85	1.68
40–49	*M*	40.74	42.24	43.72	44.35	43.37	42.45	37.14	29.00	20.86	10.25
	SD	2.79	2.75	2.56	2.30	2.41	2.43	2.00	2.79	3.67	2.26
50–58	*M*	37.89	39.92	41.15	42.43	42.41	41.77	36.94	28.53	19.07	8.50
	SD	2.62	2.87	2.59	2.29	1.94	1.91	1.89	2.95	2.91	1.47
61–68	*M*	37.48	39.63	41.32	42.31	41.63	40.87	36.22	28.17	18.17	8.06
	SD	1.35	1.54	2.22	1.82	2.15	2.22	2.48	3.53	3.35	1.95
70–78	*M*	34.89	36.99	38.56	40.32	40.68	40.46	36.63	29.83	20.38	8.09
	SD	2.39	2.52	2.54	2.59	2.67	2.75	3.00	3.94	4.73	4.36
80–89	*M*	33.09	34.89	36.34	38.12	38.70	38.29	34.57	26.97	17.67	7.32
	SD	3.06	3.63	3.71	3.88	3.80	3.46	3.42	3.52	3.73	2.21
13–89	*M*	38.86	40.64	42.13	43.01	42.41	41.70	36.81	28.89	19.69	9.20
	SD	4.33	4.26	4.28	3.83	3.31	3.13	2.77	3.29	3.62	2.39

TABLE 2

MEAN CFF'S AND *SD*'S FOR EIGHT AGE GROUPS AT 10 LIGHT:DARK RATIOS WHEN THE TEST AREA APPEARS AT THE CENTER OF A SURROUND OF .04 ML.

Age Range		Light-Time Percentage in Flicker Cycle									
		2%	5%	10%	25%	40%	50%	75%	90%	95%	98%
13–19	*M*	42.37	44.02	45.41	44.50	42.74	41.54	36.20	28.41	20.52	12.04
	SD	1.94	2.01	1.94	2.29	2.03	1.81	2.26	2.01	1.72	1.22
20–29	*M*	40.15	42.20	43.54	43.16	41.62	40.58	35.19	27.42	19.90	11.31
	SD	1.78	1.60	1.96	2.04	1.91	2.04	2.19	2.59	2.48	2.33
30–38	*M*	39.23	41.46	42.59	42.45	41.18	40.37	35.52	27.42	19.80	11.50
	SD	2.58	2.54	2.71	2.55	2.83	2.84	2.86	2.96	2.82	2.60
40–49	*M*	37.91	40.11	41.48	41.77	40.84	39.93	35.09	27.23	20.04	11.59
	SD	2.88	2.85	2.60	2.15	2.64	1.96	1.94	2.52	2.75	2.26
50–58	*M*	35.60	37.91	39.56	40.61	40.33	39.66	34.90	26.64	18.53	9.52
	SD	2.71	2.65	2.69	2.53	2.20	2.14	1.84	2.61	2.74	1.69
61–68	*M*	35.10	37.37	38.90	39.95	39.37	38.94	34.29	26.29	18.21	9.54
	SD	1.60	1.60	2.12	2.84	2.86	2.79	3.13	4.08	3.35	2.07
70–78	*M*	32.93	35.04	37.12	38.43	38.72	38.61	34.81	27.87	19.38	9.12
	SD	2.26	2.34	2.48	2.54	2.76	2.94	2.86	3.76	3.72	2.53
80–89	*M*	30.88	32.62	34.38	35.92	36.45	36.49	32.84	24.90	17.06	7.67
	SD	2.79	3.03	3.36	3.36	3.50	3.56	3.29	3.53	3.15	2.15
13–89	*M*	36.51	38.59	40.13	40.67	40.03	39.41	34.78	26.94	19.11	10.21
	SD	4.24	4.28	4.17	3.59	3.59	2.96	2.79	3.26	3.12	2.61

decreases as age increases under all LDR conditions and under both conditions of surround luminance. The *SD*'s, however, are not ordered as systematically as the means. When the youngest age group is compared to the oldest age group, the younger *O*s have a smaller *SD* at all LDR levels and at both surround luminances. This same relationship is also manifested when the 20–29 yr. group is compared to the 70–78 yr. group. The computation of Bartlett's test for homogeneity of variance, however, indicates that the variances among the eight age groups are not significantly different at any LDR level under both surround luminances. All of the 20 chi-square values derived from these data were smaller than the value required for significance at the 5% level.

The relationship between CFF and the light-time percentage of the flicker cycle for the various age groups is presented in Fig. 1. For the purpose of clarity, *O*s were divided into four groups, of approximately a 20-yr. span. In the upper part of the figure are presented the observations obtained with a surround luminance of 23.6 ml., and in the lower part are those for a surround of .04 ml. From these curves, and from the data presented in Tables 1 and 2, it becomes apparent that the change in CFF with age is a function of light-time percentage of the flicker cycle. The age differences are most marked when the proportion

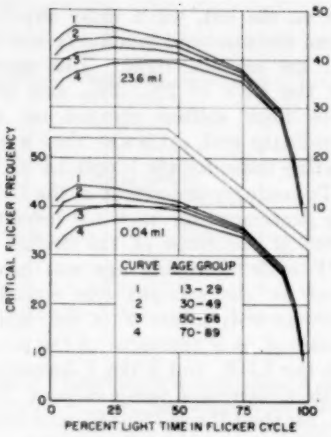


FIG. 1. Relationship between CFF and the light-time percentage of the flicker cycle for four age ranges, when surround illumination is 23.6 ml. (top), and .04 ml. (bottom).

of light in the LDR is low, and the differences become progressively less noticeable as the LP increases. Although the CFF values obtained at the higher LP's lack the regularity obtained at the lower LP levels, differences are still apparent when the extreme age groups are compared.

Inspection of the data also revealed that there is a systematic difference among the individuals in the LP level at which the maximum threshold reading is obtained. For all the individual CFF-LDR curves a comparison of the critical frequencies at the various LDR's under both conditions of surround luminance revealed that CFF increased slightly as the LP increased, and then reached a maximum between an LP of 10% and 50%. It can readily be seen from the data presented in Table 3 that the younger *O*s reach their maximum CFF value when the LP of the cycle is 10%, and as age increases, the maximum CFF progressively shifts to higher LP levels up to 50%. When the light comprises more than 50% of the cycle, CFF decreases as the LP increases for all *O*s.

TABLE 3

PERCENTAGE OF *O*s IN THE VARIOUS AGE RANGES REACHING MAXIMUM CFF AT THE DIFFERENT LIGHT-TIME PERCENTAGE LEVELS FOR THE TWO SURROUND LUMINANCES

Age Range	Percentage of Light in the Flicker Cycle							
	23.6 ml. Surround				.04 ml. Surround			
	10%	25%	40%	50%	10%	25%	40%	50%
13–29	65	35	0	0	83	17	0	0
30–49	28	66	3	3	34	55	11	0
50–68	3	61	36	0	7	61	29	3
70–89	0	18	39	43	0	18	35	47

Examination of the CFF-LDR function for all 108 *O*s for the surround luminance of 23.6 ml. (bottom of Table 1) shows that CFF increases from an initial value of 38.86 cps at an LP of 2% to a maximum CFF of 43.01 cps at an LP of 25%. Beyond 25% light-time, increases in the LP to 40%, 50%, 75%, 90%, 95%, and 98% result in a progressive decline of CFF in which the threshold values are respectively 99%, 97%, 86%, 67%, 46%, and 21% of those obtained at 25% light-time. Comparing the CFF thresholds obtained with a surround luminance of .04 ml. (bottom of Table 2), as the LP increases from 2% to 25%, CFF increases from 36.51 cps and reaches a maximum of 40.67 cps. Further increases in the percentage of light in the flicker cycle to 40%, 50%, 75%, 90%, 95%, and 98% result in a decrease in CFF in which the values are respectively 98%, 97%, 85%, 66%, 47%, and 25% of those obtained at 25% light-time.

Under the conditions of lowered surround luminance, the CFF values

decrease approximately 2 cps from 2% to 90% light-time. When the LP of the cycle is 95%, the lowered surround luminance has only a very small effect on CFF, and at 98% there is a reversal in which the threshold obtained under the surround luminance of .04 ml. is 1 cps higher than that obtained under the brighter surround luminance.

The CFF-age decrement for each LP level was represented by the best fitting straight line. This was determined by the method of least squares for the eight mean CFF values obtained with the various age groups at each LP level. The relationship between chronological age and foveal CFF is linear and negative for all the LP levels as is shown in Fig. 2. The data obtained when the surround luminance was 23.6 ml. are on the left, while those obtained when the surround luminance was .04 ml. are on the right. The curves for the LP's of 2%, 5%, and 10% have been shifted upward on the coordinate grid, otherwise they would overlap those for the longer LP's.

To compare the extent of the CFF-age decrement for each LP level the value of the slope of the decline of CFF with increasing age was calculated by the straight line equation $Y = mx + k$, where Y is the fusion threshold, m is the value of the slope, x is the LDR, and k the Y-intercept. With a surround luminance of 23.6 ml., for the LP's of 2%, 5%, and 10%, the lines are essentially parallel and the slopes are steep with m values $-.1636$, $-.1543$, and $-.1543$, respectively. As the LP beyond 10% increases to 90%, the slopes become

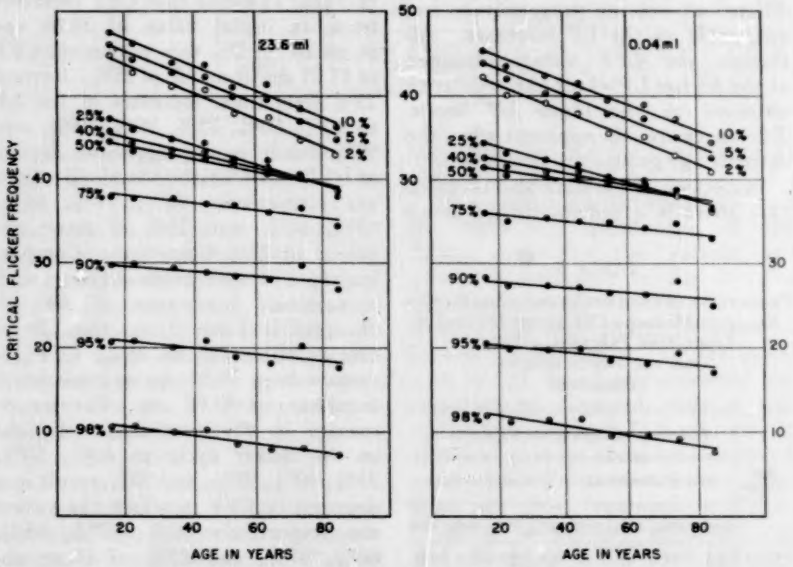


FIG. 2. Relationship between CFF and age at the various light-time percentages. Surround luminance 23.6 ml. left, .04 ml. right.

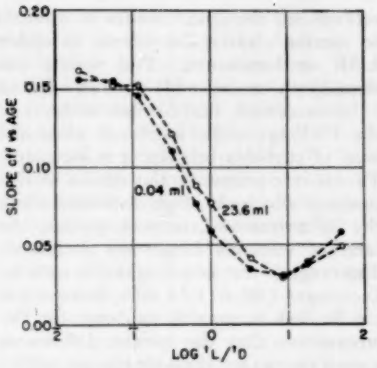


FIG. 3. Relationship between slopes of lines shown in Fig. 2 and log light-time/dark-time.

progressively less steep with m values of −.1229, −.0897, −.0786, −.0426, and −.0301 corresponding to the LP levels from 25% to 90%. At the LP levels of 95% and 98%, the progressive decline of the slopes is reversed with m values increasing to −.0385 and −.0508.

This same relationship also holds for the slopes for the various LDR's under the lower surround luminance. The decreasing slopes are −.1581, −.1570, −.1476, −.1116, −.0779, −.0601, −.0353, and −.0317, corresponding to the LP's from 2% to 90%, respectively. When the light comprises more than 90% of the cycle, there is a reversal in the rate of decay of the slopes with m values of −.0400 and −.0597 corresponding to the LP's of 95% and 98%. For both surround luminances the lines in Fig. 2 from 10% to 98% appear in descending order, showing a regular decline of CFF values. Only for the LP's of 2% and 5% is the order reversed, yielding lower CFF values than those found for 10% light-time.

The slopes of the straight lines expressing the CFF-age relationship shown in Fig. 2 are presented in Fig. 3 as functions of the logarithms of light-time/dark-time. The curve for the surround luminance of 23.6 ml. is at a slightly higher level than the curve for the surround of .04 ml. The curves are in general parallel, with the exception of the extreme LDR's. In each curve three segments can be distinguished. The middle segment is almost straight and has a steep slope indicating an inverse relationship between the slopes of the lines in Fig. 2 and the proportion of light in the flicker cycle. The portions of Fig. 3 which deviate from the straight line are on the upper left (2% and 5% light-time), and at the lower right (90%, 95%, and 98% light-time). The slopes for 2% and 5% light-time are not much higher than for 10% light-time, and for this reason the points drop off to the left. With 95% and 98% light-time the slopes do not decrease in the same way the slopes corresponding to those between 10% and 90% light-time decrease. Beyond an LP of 90%, further increases in the light-time resulted in an increase of the CFF-age slope. This was in contrast to the relationship observed from 2% to 90%.

DISCUSSION

Among the conditions known to influence CFF is the relative light proportion of the flicker cycle, but the CFF-LDR function has not been well defined. The only consistent relationship observed by the majority of the investigators is that, beyond an LP of 50%, CFF decreases as light-time increases. Discrepancies in the experimental studies appear primarily in (*a*) the relationship between CFF and LDR with light pulses shorter than 50% of the cycle, (*b*) on the specific rate of decline of CFF beyond an LP of 50%, and (*c*) on the particular LDR where CFF reaches a maximum.

The CFF-LDR function obtained in this study is not intended to resolve the diversified results of the previous studies (1,3,7), particularly since Landis (9) has demonstrated that the interaction of various physical properties of the stimulus, such as intensity and area, results not only in a change of the position of the curves but in the direction of their slopes. The significant aspect of the results obtained with the flicker fusion apparatus in this experiment, other than the form of the function, is the indication that the chronological age of *O* is another interacting variable in the determination of the CFF-LDR relationship. The brightness of the surround, however, does not appear to add to the interacting factors since the shapes and slopes of the two curves were the same. The effect of the two surround luminances resulted only in a shift in the position of the curves on the frequency axis.

The analysis of the CFF-age curves as a function of LDR demonstrated that the variation of the light-time of the flicker cycle produced changes in *O*'s fusion threshold. Of greater significance, however, is the fact that the sensitivity of CFF as an indicator of the visual aging process was enhanced under the conditions of stimulation provided by the lower light-times. The total drop in CFF as a function of age is largest and about equal for the light-time percentages of 2%, 5%, and 10% for both conditions of surround luminance. Between 10% and 75% light-times the decrement in foveal CFF is inversely proportional to the log light-time/dark-time. Further increases in the light-time to 95% and 98% resulted in an inflection of the Slope-LDR curve indicating that the slope of the line expressing the CFF-age relationship increased slightly.

It should be kept in mind that this experiment was conducted using the uncompensated luminance method. The increase in luminance accompanying the increasing increments in the light percentage of the cycle makes it difficult to ascribe clearly the effects to either LDR or luminance. The results obtained by Coppinger (4) with an LDR of 1:1 demonstrate that the rate of decay of the CFF-age slope increases when the level of stimulus brightness is increased. This is in contrast to the results of this study in which the slope decreased when the light-time and, in consequence, the level of stimulus brightness increased. The range of test area brightness used by Coppinger (.09 to 1.74 ml.), however, is too limited to provide evidence for the proposition that the various differences among the age groups is clearly not attributable to the changing luminance levels. It is also possible that there may be an interaction between luminance and LDR. Thus, the lower levels of test area luminance, together with the lower LP levels, may have produced the effect such that any increase in luminance could have abolished the effect. Clearly, the relative contribution of each of these variables should be determined.

In tests of sensitivity of an *O*, his verbal or motor response to a visual stimulus is obviously the terminal reaction to a long chain of events. Thus, if response thresholds vary with age, it can be safely assumed that the effectiveness of the stimulus at any or several points from the receptor to the effector has been altered and may involve conditions which are ophthalmological, cardiovascular, and/or neural in nature. Without attempting to speculate in terms of the formulation of any specific causative factor, Coppinger maintains that "many of the structural and functional changes associated with aging are qualitatively similar to the diseased and artificially induced conditions which lower CFF in younger persons" (4). Therefore, in line with this reasoning, and in view of the differential effect of LDR as noted in this study, it seems reasonable to anticipate the likely hypothesis that the effect may be manifested in situations other than the aging process. A review of the literature reveals that

possible evidence to support this proposition is provided by the findings of Davis (5). In her investigation, fusion thresholds were determined using both an episcotister (50:50 LDR), and a Strobotac (1:99.9 LDR) at three levels of stimulus brightness: .195, .75, and 3.0 ft.-c. The thresholds provided by the Strobotac with the shorter LP proved to be more sensitive indicators to the physiologic change induced by chemical agents. This was true at all test area luminance levels.

The literature is replete with statements concerning the effectiveness or ineffectiveness of CFF as an indicator measure, but very few investigators have been concerned with the determination of the stimulus conditions which enhance the use of CFF as an indicator response in the assessment of the neurophysiological condition of *O*. Aside from purposes of standardization, there is no a priori rationale or any experimental findings which suggest the continued use of the conventional 50:50 LDR under one level of luminance in flicker research. This observation has also been made by Winchell and Simonson who state that "the importance of the standardization of LDR would depend on the magnitude of changes produced by its variation, and on the sensitivity of the CFF at different LDR's in response to physiological stress" (16, p. 188).

Summary

Binocular critical flicker frequencies (CFF's) were determined for 108 *O*s ranging in age from 13 to 89 yr. The light pulses were in the form of square waves activating a glow modulator tube. The ratio of light-time/dark-time in the flicker cycle was varied systematically from 2/98 to 98/2. The uncompensated luminance procedure was used. The test field consisted of a circular test area of .6° subtense shown at the center of a circular 17° surround. Tests were made at surround luminances of 23.6 and .04 ml.

The major results were: (*a*) The relationship between CFF and age is linear and negative for all LDR levels under both conditions of surround luminance. (*b*) The decrement of CFF with age is a function of the light-time percentage of the flicker cycle. The differentiation between age groups is more pronounced at the lower light-time percentages. (*c*) The light-time percentage at which the maximum CFF occurs is a function of the chronological age of *O*. (*d*) The change of the surround luminance resulted in a shift of the position of the CFF-LDR curves, but the shape and slope of the curves were not altered.

On the basis of the results it was suggested that the use of lower levels of light-time in the flicker cycle, rather than the conventional 50:50 LDR, would enhance the sensitivity of CFF as an indicator measure.

References

1. Bartley, S. H. Neural determination of critical flicker frequency. *J. exp. Psychol.*, 1937, **21**, 678–686.
2. Brozek, J., & Keys, A. Changes in flicker-fusion frequency with age. *J. consult. Psychol.*, 1945, **9**, 87–90.
3. Cobb, P. W. Dependence of flicker on the dark-light ratio of the stimulus cycle. *J. opt. Soc. Amer.*, 1934, **24**, 107–113.
4. Coppinger, N. W. Relationship between critical flicker frequency and chronological age for varying levels of stimulus brightness. Unpublished Ph.D. dissertation, Tulane Univer., 1951.
5. Davis, R. J. A comparison of the stability of the measures of critical flicker-fusion made at two different light-dark ratios as provided by the episcotister and the Strobotac. Unpublished M.A. thesis, Columbia Univer., 1952.
6. Fisher, M. B. The relation of age to the dependency of critical flicker frequency on light-dark ratio. *Amer. Psychologist*, 1955, **10**, 333. (Abstract)
7. Ives, H. E. Critical frequency relations in scotopic vision. *J. opt. Soc. Amer.*, 1922, **6**, 254–268.
8. Landis, C. Something about flicker fusion. *Sci. Mon., N. Y.*, 1951, **73**, 308–314.
9. Landis, C. Determinants of critical flicker-fusion thresholds. *Physiol. Rev.*, 1954, **34**, 259–286.
10. Miles, P. W. Flicker fusion fields. I. The effect of age and pupil size. *Amer. J. Ophthal.*, 1950, **33**, 769–773.
11. Miles, P. W. Flicker fusion fields. II. Technique and interpretation. *Amer. J. Ophthal.*, 1950, **33**, 1069–1077.
12. Misiak, H. Age and sex differences in critical flicker frequency. *J. exp. Psychol.*, 1947, **37**, 318–332.

13. Simonson, E., & Brozek, J. Flicker fusion frequency: Background and applications. *Physiol. Rev.*, 1952, **32**, 349–378.
14. Simonson, E., Enzer, N., & Blankstein, S. S. The influence of age on the fusion frequency of flicker. *J. exp. Psychol.*, 1941, **29**, 252–255.
15. Weekers, R., & Roussel, F. La mesure de la fréquence de fusion en clinique. *Documenta Ophthalmol.*, 1948, **2**, 130–192.
16. Winchell, P., & Simonson, E. Effect of the light:dark ratio on the fusion frequency of flicker. *J. appl. Physiol.*, 1951, **4**, 188–192.

(Received for early publication
July 11, 1958)

ZEITSCHRIFT FÜR EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Journal of the German Psychological Association

Publisher in Germany

PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Publisher for Austria	Publisher for Switzerland	Editorial Office
PROF. DR. H. ROHRACHER, Vienna	PROF. DR. R. MEILI, Bern	DR. C. J. HOGREFE, Göttingen

Vol. 5, No. 2

REINHARD TAUSCH. **The Nature of the Relationship Between the Therapist and the Client and the verbal role of the therapist in Nondirective Psychotherapy**

WERNER FRÖHLICH. **The Aubert-Foerster Phenomenon and the constriction of the visual field as a Result of a Moving Background**

HANS-JÜRGEN PFISTNER. **On the Contents of the Physiognomic and Pathognomic Expressive Phenomena**

ANNELIESE KRANTZ-GROSS and HANS-JOACHIM HAASE. **Writer's Cramp as Compared to Other Psychomotor Phenomena in Endogenous Depressives and Normals**

FRANZ HILLINGER. **Introversion and "Set" Among Siblings**

ANTON ZRZAVY. **A New Method of Validity Control in Personality Tests (Testdoppelganger-Methode)**

JURGEN SCHMIDT. **First Steps in the Development of Methods for Psychological Tests of the Blind**

KLAUS LÖHN AND WERNER TRAXEL. **On an Electronic Sound Key**

All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per issue.
Subscription: DM 64.—per volume.

VERLAG FÜR PSYCHOLOGIE • DR. C. J. HOGREFE • GÖTTINGEN
Brentanoweg 10, Germany

AMERICA'S PSYCHOLOGISTS

A Survey of a Growing Profession

KENNETH E. CLARK

Associate Professor of Psychology

University of Minnesota

A report on a three-year study of American Psychologists, Sponsored by the Policy and Planning Board of the American Psychological Association and the National Science Foundation. The volume presents a view of a growing profession in the mid-twentieth century by describing the people active in psychology. $1.00

Order from

American Psychological Association, Inc.

1333 Sixteenth Street, N.W.

Washington 6, D.C.